

### Das unterschätzte Potenzial hoher Fallzahlen - Stärken und Limitierungen des Mikrozensus am Beispiel von Fertilitätsanalysen

Bujard, Martin; Dorbritz, Jürgen; Herter-Eschweiler, Robert; Lux, Linda

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

Verlag Barbara Budrich

#### Empfohlene Zitierung / Suggested Citation:

Bujard, M., Dorbritz, J., Herter-Eschweiler, R., & Lux, L. (2015). Das unterschätzte Potenzial hoher Fallzahlen - Stärken und Limitierungen des Mikrozensus am Beispiel von Fertilitätsanalysen. *Zeitschrift für Familienforschung*, 27(3), 343-372. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-456456>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-sa/4.0>

*Martin Bujard, Jürgen Dorbritz, Robert Herter-Eschweiler & Linda Lux*

## Das unterschätzte Potenzial hoher Fallzahlen – Stärken und Limitierungen des Mikrozensus am Beispiel von Fertilitätsanalysen

### The unrecognised potential of large N – Strengths and limitations of the German Microcensus exemplified by fertility research

#### **Zusammenfassung:**

Aufgrund der hohen Fallzahlen (688.931 Personen in 2012) ist das Potenzial des Mikrozensus (MZ) für demografische und familiensoziologische Analysen groß, zudem inhaltlich durch die Erhebung der Geburten von Frauen noch gestiegen. Der Beitrag hat zum Ziel, methodische Hintergründe, Limitierungen und Potenziale des MZ am Beispiel von Fertilität systematisch aufzuzeigen. Die Analyse von Antwortverweigerungen zeigt, dass die Art der Befragung einen starken Einfluss hat, deren Verzerrungen durch Imputationen deutlich reduziert werden konnten. Die Limitierungen sind v.a. das Querschnittsdesign sowie die fehlenden Daten zur Kinderzahl von Männern und zur Binnenmobilität. Die hohen Fallzahlen des MZ ermöglichen tiefenscharfe Differenzierungen bezüglich Sozialstruktur, Paritäten und Zeitverläufe, was anhand von Konfidenzintervallen diskutiert wird.

Drei Forschungsdesigns, die die Analysepotenziale nutzen, werden vorgestellt: Dendrogrammanalysen ermöglichen gruppenspezifische Differenzierungen nach fünf Ebenen und zeigen so ein detailliertes Muster der Kinderlosigkeit. Der Verlauf von Paritätsmustern im Kohortenvergleich verdeutlicht die je nach beruflicher Bildung unterschiedlichen Verläufe. Die Analysen zur altersspezifischen Häufigkeit von Erstgeburten ermöglichen, Trends von nachholenden Geburten für verschiedene Bildungsgruppen zu identifizieren.

**Schlagwörter:** Mikrozensus, quantitative Erhebung, Imputationen, Non-Response, Fertilität, Kinderlosigkeit, Sozialstruktur

#### **Abstract:**

Due to a large number of cases (688.931 persons in 2012) the German Microcensus has a huge potential for analyses in demography and family sociology. Content is further improved by the new voluntary question on biological children of women. This paper aims to systematically assess the methodological background as well as limitations and potentials of the Microcensus using the example of fertility data. The analyses of non-response show that the way of questioning has a strong impact. The related bias could, however, be reduced by imputations. The main limitations are the cross-sectional design and missing data on the children of men and internal migration. The potential of the high case numbers of the Microcensus is systematically discussed by analyzing confidence intervals: It allows for in-dept differentiations for analyzing social structure, birth parities and time patterns.

Finally, we discuss three research designs which are able to use this potential: Dendrogram analyses allow group differentiations for up to five levels and thus reveal detailed patterns of childlessness. Parity patterns by cohorts illustrate differences by occupational education. Third, the age-specific analyses of first births among several cohorts allow for identifying trends of recuperation which differ considerable between educational groups.

**Key words:** Microcensus, quantitative survey, imputations, non-response, fertility, childlessness, social structure

## 1. Einleitung

Der Mikrozensus (MZ) unterscheidet sich fundamental von sozialwissenschaftlichen Erhebungen wie dem Sozio-oekonomischen Panel (SOEP), pairfam, NEPS, FiD oder GGS: Zum einen sind die Fallzahlen des MZ 2012 erheblich höher, zum anderen ist er eine Querschnittserhebung. Während das fehlende Paneldesign<sup>1</sup> erhebliche methodische und kausale Einschränkungen beinhaltet, ermöglichen die hohen Fallzahlen strukturelle Muster und Trends differenziert und stabil zu erfassen. Der MZ 2012 beinhaltet 688.931 befragte Personen bezogen auf die Wohnbevölkerung. Betrachtet man nur die Personen am Hauptwohnsitz – auf die sich in der Regel die Aussagen beziehen – wurden 680.563 Personen befragt. Da die hochgerechnete Bevölkerung am Hauptwohnsitz im MZ 2012 81,9 Mill. beträgt, liegt der Anteil der Befragten bei 0,83%.<sup>2</sup> Die Fallzahlen für Frauen im Alter von 15 bis 65 Jahren betragen pro Altersjahr zwischen 3.118 (17 Jahre) und 5.949 (48 Jahre). *Pro Altersjahr* der Befragten hat der MZ folglich etwa die Größenordnung an Fällen, die andere Datensätze *insgesamt* von Frauen haben.

Die Potenziale des MZ liegen in der differenzierten Erfassung von Trends und Strukturen hinsichtlich von drei Dimensionen:

1. sozialstrukturelle Differenzierung,
2. paritätsspezifische Differenzierung und
3. zeitliche Differenzierung nach einzelnen Kohorten und Altersjahren.

Der MZ ermöglicht eine Kombination von Differenzierungen aller drei Dimensionen – beispielsweise die Entwicklung der Kinderlosigkeit bei Akademikerinnen zwischen einzelnen Geburtsjahrgängen oder den Anteil kinderreicher Frauen in ländlichen Regionen im Vergleich zwischen einzelnen Altersjahren. Er ermöglicht auch Analysen von spezifischen Merkmalen differenziert nach Subgruppen – beispielsweise der gleichzeitigen Differenzierung nach Bildung, Urbanisierungsgrad, Arbeitszeit, Migrationshintergrund und Partnerschaft. Zudem lassen sich präzise zeitliche Verläufe und Trends identifizieren wie die Entwicklung von Kinderlosigkeit nach Kohorten. Durch die Kombination mehrerer MZ lassen sich sogar Muster von Altersjahren und Kohorten kombinieren und dadurch präzise gesellschaftliche Veränderungsprozesse identifizieren.

Für Fertilitätsdaten in Deutschland (zum Überblick: Kreyenfeld et al. 2011) spielt der MZ eine zentrale Rolle. Warum ist das so? Während in vielen anderen Ländern die Geburtenstatistik Angaben zur ordnungsspezifischen Geburtenfolge bereitstellt, wurde dies in Deutschland erst im Jahr 2009 begonnen. Da jedoch damit erst für Frauen ab dem Jahrgang 1994 die Kinderlosigkeit gemessen werden kann, wird die Kinderlosigkeit für 40-jährige Frauen erst 2034 und die endgültige mit 45 bzw. 49 Jahren entsprechend später vorliegen (vgl. Kreyenfeld/Konietzka 2013). Zudem ermöglicht die Geburtenstatistik nur wenige sozialstrukturelle Differenzierungen. Allerdings eignet sie sich für Analysen der altersspezifischen Fertilität und des Geburtenaufschubs (Pötzsch 2013). Sozialwissenschaftliche Datensätze wie das SOEP, GGS oder pairfam sind, abgesehen von der Verzer-

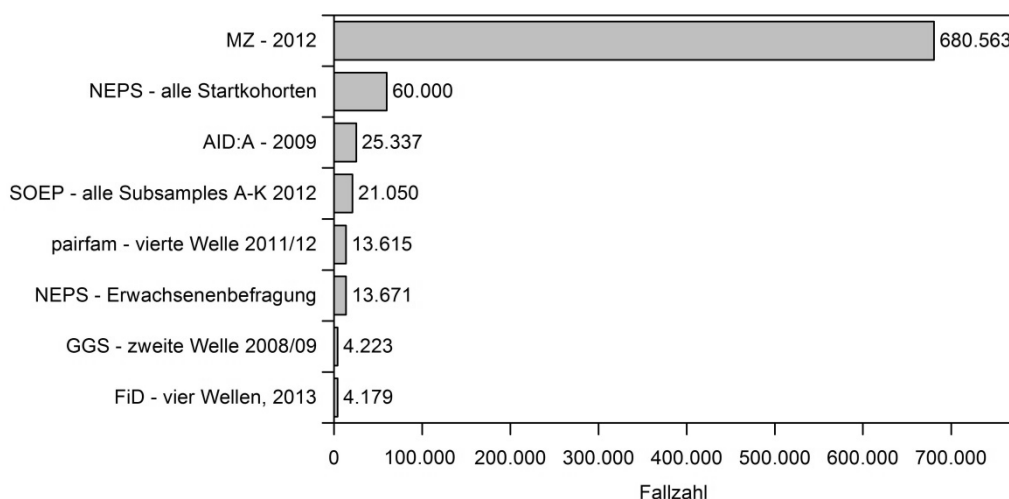
---

1 Eine Ausnahme ist der Vier-Jahres-Panel, der jedoch nicht für die Geburtenfrage vorliegt. Zu den Potenzialen des MZ-Panel siehe: Kreyenfeld et al. (2009).

2 Diese Werte sind nicht an die Bevölkerungsfortschreibung auf Basis des Zensus 2011 angepasst.

rung durch Antwortverweigerer, v.a. aufgrund ihrer Fallzahl nur begrenzt geeignet, um die Entwicklung von Kinderlosigkeit über einzelne Jahrgänge (und Bevölkerungsgruppen) exakt zu quantifizieren. Die Fallzahlen des MZ 2012 sind um das 11- bis 160-fache höher (Abb. 1).

Abbildung 1: Fallzahlen sozialwissenschaftlicher Datensätze in Deutschland



Quelle: DIW/SOEP 2013; Dreschmitt/Naderi 2015; eigene Recherche; eigene Darstellung.

Bis vor wenigen Jahren war die Datenlage zur Fertilität im MZ problematisch (Deutscher Bundestag 2007; Kreyenfeld/Konietzka 2013; Statistisches Bundesamt 2006). In den Mikrozensusen seit 1972 wurde nur erhoben, ob und wie viele Kinder im Haushalt leben.<sup>3</sup> Dadurch ließen sich die durchschnittliche Kinderzahl und der Anteil kinderloser Frauen nicht exakt berechnen. Erst durch das „Gesetz zur Änderung des Mikrozensusgesetzes 2005“ vom 30.10.2007 wurde dieser Datenmangel behoben – allerdings nicht für Männer (vgl. Stock et al. 2012). Ab 2008 werden alle vier Jahre Frauen im Alter zwischen 15 und 75 Jahren nach der Anzahl der leiblichen Kinder im MZ gefragt. Dies stellt eine fundamentale Verbesserung für die demografische und sozialwissenschaftliche Forschung dar.

Der Kontrast zwischen der Bedeutung des Phänomens Kinderlosigkeit und der lange Zeit fehlenden sicheren Datengrundlage ist beachtlich. Die Bundesrepublik Deutschland ist länger als alle anderen Länder der Welt ein Low-Fertility-Land, die Total Fertility Rate (TFR) liegt seit 1975 unter 1,5. Deutschland hat einen der höchsten Anteile endgültiger Kinderlosigkeit weltweit (vgl. Kohler et al. 2002, Kreyenfeld/Konietzka 2013; OECD 2015). Über viele Jahre hinweg existierten heterogene und teilweise stark überhöhte Zahlen zur Kinderlosigkeit in Deutschland. Für die Jahrgänge Mitte der 1960er Jahre wurden sogar Kinderlosenanteile von etwa 30% geschätzt (Birg 2003, vgl. Kohler et al. 2002), während Mietinen et al. (2015) als aktuellsten Wert für endgültige Kinderlosigkeit basie-

3 Bereits 1962 und 1963 wurde, allerdings auf die Ehe bezogen, die Frage gestellt: „Wieviele Kinder – auch wenn sie nicht zum Haushalt gehören – wurden in der jetzigen Ehe lebend geboren?“ Allerdings wurden MZ-Daten vor 1971 im Statistischen Bundesamt gelöscht.

rend auf GGS-Daten für die Kohorten 1960-64 14,2% angeben. Wie die Beispiele zeigen, ist das Spektrum der Schätzungen zur Kinderlosigkeit in Deutschland groß, was auf eine jahrzehntelange unzureichende Datenbasis zurückzuführen ist. Tatsächlich liegt die Kinderlosigkeit der Kohorten 1960-64 bei 18,7% und 1965-69 bei 20,8% und 1971 bei 21,9%.<sup>4</sup>

Seit Kurzem liegen also für den MZ Fertilitätsinformationen vor. Zudem ist die 2012er Befragung gegenüber der Erstbefragung 2008 deutlich verbessert worden (siehe 3.3). Genauere Analysen und neue Erkenntnisse sind nun möglich. Während die Rahmen- und Daten zur Kinderlosigkeit (Statistisches Bundesamt 2009a, 2013a) und erste bivariate Auswertungen (u.a. Dorbritz/Ruckdeschel 2013) bereits publiziert sind, sind viele Potenziale noch nicht gehoben.

Der Beitrag hat zum Ziel, die Potenziale und Limitierungen des MZ am Beispiel der Fertilitätsforschung aufzuzeigen. Die Ergebnisse und die verwendeten Forschungsdesigns lassen sich aber auch auf MZ-Analysen zu anderen Themen übertragen. Dazu werden zunächst anhand von Auswertungen der MZ 1972-2013 Möglichkeiten gezeigt, die Kinderlosigkeit mit den Informationen zur Kinderzahl im Haushalt zu schätzen (Kapitel 2). In Kapitel 3 werden die Grenzen, Potenziale und methodischen Herausforderungen des MZ diskutiert. Dabei werden am Beispiel von Kinderlosigkeit Befunde des MZ in Hinblick auf Fallzahlen, Abweichungen und Konfidenzintervalle mit dem SOEP verglichen, da das SOEP einen der einflussreichsten Mikrodatensätze in Deutschland darstellt. Auch werden kritisch die methodischen Aspekte von befragungsbedingten Verzerrungen und Imputationen diskutiert. In Kapitel 4 werden dann drei Beispiele für Forschungsdesigns angewendet, die die Differenzierungsmöglichkeiten anhand der drei Dimensionen Sozialstruktur, Parität und Jahrgänge aufzeigen.

## 2. Anteil kinderloser Frauen in Mikrozensus vor 2008

### 2.1 *Überhöhte Schätzungen des Anteils Kinderloser und neue Einsichten durch genauere Messung*

Vor den Mikrozensus 2008 und 2012 waren Bevölkerungswissenschaftler angewiesen auf Schätzungen, Surveydaten mit relativ niedrigen Fallzahlen, Auszählungen des MZ nach Altersgruppen und der Zahl der im Haushalt lebenden Kinder oder die Geburtenstatistik, in der die Ordnungsfolge der lebendgeborenen Kinder nur in der jeweils bestehenden Ehe gezählt wurde.

Bei den Schätzungen soll vor allem auf die Arbeiten von Birg und Flöthmann (1996) sowie Sobotka (2005) hingewiesen werden. Birg und Flöthmann haben Daten der Perinatalstatistik der Ärztekammer Westfalen-Lippe ausgewertet und davon ausgehend die Kinderlosigkeit für Westdeutschland geschätzt. In die Analyse einbezogen waren die Ge-

---

4 Eigene Berechnungen auf Basis des MZ 2012. Die Werte für Westdeutschland sind mit 20,3% für 1960-64, 21,9% für 1965-69 bzw. 22,6% für 1971 höher, in Ostdeutschland mit 12,0%, 15,8% bzw. 18,8% niedriger.

burtsjahrgänge 1940-1970. Für den Geburtsjahrgang 1940 ist ein Wert von 10,6% ermittelt worden (MZ 2012: 12,8%). Für den Jahrgang 1970 ist die Kinderlosigkeit mit 32,1% etwa um 10 Prozentpunkte überhöht ausgewiesen worden (MZ 2012: 21,5%). Auch in späteren Publikationen hat Birg (u.a. 2003) Schätzungen zur endgültigen Kinderlosigkeit der Mitte der 1960er Jahre geborenen Frauen von 30% popularisiert. Sobotka hat für seine Schätzungen EUROSTAT-Daten aus dem Jahr 2003 herangezogen und für die Kohorten 1960-1975 in Westdeutschland zwei Szenarien berechnet. Der Anteil kinderloser Frauen wird für den Jahrgang 1960 zwischen 19,3 und 19,7% angegeben und entspricht den Werten des MZ 2012 für Westdeutschland von 19,7%. Für den Jahrgang 1975 liegen Sobotkas Angaben in einem Bereich zwischen 23,0 und 27,7%. Verdienst dieser Schätzungen war es, früh auf das hohe Niveau der Kinderlosigkeit in Westdeutschland hingewiesen und damit für eine größere politische Aufmerksamkeit gesorgt zu haben.

Die Auswertungen der Mikrozensen vor 2008 haben ebenfalls auf die steigende Kinderlosigkeit hingewiesen, sie aber insbesondere bei den Hochqualifizierten zum Teil deutlich überschätzt. Die Kinderlosigkeit von Akademikerinnen ist teilweise mit ca. 40% angegeben worden (vgl. Duschek/Wirth 2005; Statistisches Bundesamt 2006). In den Geburtsjahrgängen, die in der zweiten Hälfte der 1960er Jahre geboren wurden, liegen die tatsächlichen Werte zwischen 30 und 31%. Die Überschätzungen sind dadurch zustande gekommen, dass der Anteil der Frauen ohne Kinder im Haushalt bei den Akademikerinnen in der Altersgruppe 35-39 Jahre 40% betragen hat. Dies führte gerade in der Gruppe der Hochqualifizierten, die ihre Kinder auch aufgrund der längeren Ausbildungswege erst in einem späteren Lebensabschnitt bekommen, zu Überschätzungen. Kritisch betrachtet wurde die sogenannte 40%-These durch Wirth (2006), Scharein und Unger (2005) sowie Schmitt und Wagner (2006).

Im Jahr 2004 wurde diese Problematik auf einem Workshop des Zentrums für Umfragen, Methoden und Analysen (ZUMA) thematisiert. Eine dazu durchgeführte Studie erbrachte erheblich niedrigere Anteile kinderloser Akademikerinnen. Wirth (2006) sowie Wirth und Dümmler (2004, 2005) stellen fest, dass die Altersabgrenzung von 35-39 Jahren bei Frauen mit Universitätsabschluss aufgrund des Geburtenaufschubs suboptimal ist und schlagen 39-42 Jahre vor (Wirth/Dümmler 2005: 332). Bei Analysen nur für die 39- bis 40-jährigen Akademikerinnen „liegt die Kinderlosenquote für die Geburtskohorten 1953 bis 1960 zwischen 28 und 34 Prozent“ (Wirth 2006: 3). Zu ähnlichen Schlussfolgerungen gelangen auch Scharein und Unger (2005: 6ff.), die vorschlugen, statt der Altersgruppe 35-39 Jahre die Altersgruppe 39-43 Jahre zu verwenden. Basierend auf dem MZ 2003 zeigen sie, dass die Kinderlosigkeit von Akademikerinnen bei 35- bis 39-Jährigen 39% und bei 39- bis 43-Jährigen etwa 30% beträgt (Scharein/Unger 2005: 9). In einem Beitrag von Schmitt und Wagner (2006) sind die Anteile kinderloser Frauen anhand der SOEP-Daten deutlich niedriger angegeben. Akademikerinnen (Absolventinnen von Fachhochschule oder Universität) der Jahrgänge 1951-1965 bleiben demnach zu 23% dauerhaft kinderlos, Universitätsabsolventinnen zu 30%. Schmitt und Wagner (2006: 315) betonen: „Dieser Wert liegt zwar immer noch über dem Durchschnitt, aber deutlich unterhalb der vielzitierten 40%-Marke, die sich allein auf Frauen mit universitärem Abschluss bezieht.“

## 2.2 Methode zur Schätzung der Kinderlosigkeit nach Informationen zu Kindern im Haushalt

Betrachtet man die Kinderlosigkeit von Frauen nach ihrem Alter auf Basis der Informationen zu den Kindern im Haushalt, also dem einzigen Zugang zu dieser Frage in den Mikrozensen zwischen 1972-2007, ergibt sich eine u-förmige Kurve mit dem niedrigsten Kinderlosenanteil im Alter von etwa 40 Jahren (siehe ausführlich: Bujard/Lück *in diesem Band*). Bei jüngeren Frauen ist die Fertilitätsbiografie noch nicht abgeschlossen, während bei älteren Frauen zunehmend mehr Kinder bereits aus dem Haushalt ausgezogen sind. Im Alter von 39-41 Jahren sind die Werte die beste Annäherung an die reale Kinderlosenquote. Im Alter von 41 Jahren liegt die Kinderlosigkeit bereits nahe an der endgültigen Kinderlosigkeit eines jeweiligen Jahrgangs. Im MZ 2012 liegt die reale Kinderlosigkeit bei 21,5%, nach der Haushaltsfrage ist sie mit 25,6% rund vier Prozentpunkte höher. Ab 44 Jahren ist sie durch den Auszug der Kinder derart verzerrt, dass sie nicht interpretierbar ist.

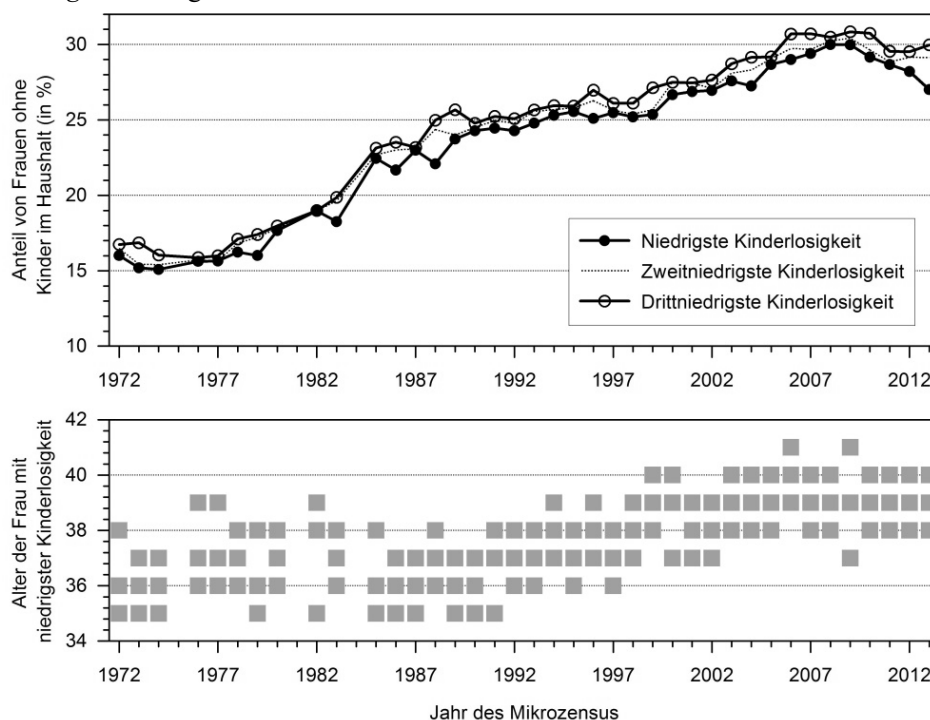
Nachdem die Erfassung der Kinderlosigkeit nach der Kinderzahl im Haushalt dargestellt wurde, wird im Folgenden der Frage nachgegangen, welche Erkenntnisse man bei einer adäquaten Nutzung dieser Information generieren kann – und wo die Grenzen liegen. Wenn man die einzelnen Mikrozensen nach dem Anteil kinderloser Frauen im Haushalt nicht – wie üblich – für Fünf-Jahres-Altersgruppen betrachtet, sondern für die Altersjahre von 34-44 Jahren einzeln, erhält man differenziertere Zahlen (vgl. Wirth/Dümmeler 2004).<sup>5</sup> Die obere Grafik von Abbildung 2 zeigt für sämtliche Mikrozensen von 1972-2013<sup>6</sup> die drei Altersjahre mit dem niedrigsten Anteil von Frauen ohne Kinder im Haushalt. Der jeweils niedrigste Wert steigt von 15-16% in den 1970er MZ auf rund 25% in den 1990er Jahren (zwischen 24,3% und 25,6%) und auf 26,7-30% in den Nuller Jahren. Seit dem höchsten Wert des MZ 2008 ist dieser Wert bis zum MZ 2013 wieder rückläufig. Die Grafik verdeutlicht, dass der zweit- und drittniedrigste Wert in allen MZ relativ nahe an dem niedrigsten liegen.

Das Alter der Frauen mit der niedrigsten Kinderlosigkeit im Haushalt unterscheidet sich zwischen den Mikrozensen. Der Trend ist aber deutlich, dass dieses Alter in den letzten vier Jahrzehnten angestiegen ist. Es war im MZ 1973 bei 35 Jahren und bei den MZ 2006 und 2013 bei 40 Jahren. Da das Alter mit der niedrigsten Kinderlosigkeit im Haushalt im Zeitverlauf stark schwankt, lässt sich kein ideales Alter für die Abschätzung der Kinderlosigkeit identifizieren. Für jeden MZ ist ein Blick auf mehrere Altersjahre notwendig. Um aber den Trend zu verdeutlichen, wurden im unteren Teil der Abbildung 2 die drei Altersjahre für die drei Spitzenwerte angegeben.

5 Hier wurde der vollständige Originalfile verwendet. Bei den Scientific Use Files (SUF) ist die Fallzahl etwas geringer, die Argumentation gilt hier aber analog (vgl. Schimpl-Neimanns 2011).

6 Für die Jahre 1975, 1983 und 1984 liegen keine Mikrozensen vor. Für 1983 wurden Werte der Arbeitskräfteerhebung verwendet. Die Werte für 1981 wurden aufgrund unplausibler Ausreißer nicht verwendet.

Abbildung 2: Vergleich des Kinderlosenanteils von 1972-2013 in Westdeutschland



Quelle: Eigene Berechnung und Darstellung, Datenbasis Mikrozensus 1972-2013.

Lesebeispiel für das Jahr des MZ 1972: Die Punkte im oberen Grafen zeigen, dass 1972 der niedrigste Anteil von Frauen ohne Kinder im Haushalt bei 16,0, der zweit- und drittniedrigste bei 16,5 und 16,8% lagen. Der untere Teil zeigt, dass diese Werte bei Frauen im Alter von 35, 36 und 38 Jahren gemessen wurden.

Die Altersjahre entsprechen den Kohorten 1934-1972. Wenn man den Wert für die Kinderlosigkeit im Haushalt bei den 38-jährigen Frauen im MZ 1972 von 16,0% mit der endgültigen Kinderlosigkeit nach der Geburtenfrage des Jahrgangs 1934 von 11,9% vergleicht, ergibt sich eine Überschätzung von 4,1 Prozentpunkten. Diese Überschätzung liegt ab den MZ 1985 meistens etwas höher. Diese MZ-Werte zur Kinderlosigkeit auf Haushaltsebene dürfen also keinesfalls als Schätzungen zur endgültigen Kinderlosigkeit interpretiert werden. Sie geben aber Obergrenzen zur Kinderlosigkeit an, die deutlich unter den zwischen 1995 und 2005 diskutierten Werten liegen. Wohlgedenkt: Die Werte für die jeweiligen MZ konnten damals bereits mit wenig Aufwand berechnet werden. Im Jahr 2000 konnte man mit dem MZ 1999 sehen, dass die Kinderlosigkeitsrate der Enddreißiger deutlich unter 25,3% liegen muss.

Zu weitaus besseren Annäherungen – bzw. zu realistischen Schätzungen – kommt man, wenn man von der Kinderlosigkeit im Haushalt der 35-Jährigen den Anteil der Frauen abzieht, die ab dem 36. Geburtstag ihr erstes Kind zur Welt bringen (vgl. Schneider 1996). Nehmen wir als Beispiel die Kohorte 1964: Kinderlosigkeit mit 35 Jahren 27,9%. Die Erstgeburten für diesen Jahrgang betragen mit 36 Jahren 1,5%, danach 1,5%,



0,9%, 0,8%, 0,4%, 0,4% und mit 42 Jahren 0,3% (eigene Berechnungen aus MZ 2000-2006). Die Kinderlosigkeit reduziert sich demnach noch um 5,8 Prozentpunkte. Der Schätzwert liegt hier bei 22,1% – ziemlich nah an der CTFR des Jahrgangs 1964 für Westdeutschland gemäß MZ 2012 von 21,3%.

Dieses Beispiel lässt sich auch für die Berechnung bei Akademikerinnen übertragen.<sup>7</sup> Allerdings ist der Vergleich der Kinderlosigkeit nach der Kinderzahl im Haushalt in früheren MZ und dem MZ 2012 durch Zu- und Abwanderung sowie Todesfälle verzerrt. Dadurch sind die befragten 37-Jährigen des Jahres 1974 nicht identisch mit der 2012 befragten Kohorte 1937. Dazu kommen spezifische Verzerrungen bei der Geburtenfrage in den MZ 2008 und 2012 (siehe 3.3).

Die oben aufgeführten Beispiele zeigen, welche Potenziale der MZ ohne Geburtenfrage bereits hat und dass manche überhöhten Werte in früherer Literatur und Medien nicht nur den Unzulänglichkeiten des MZ, sondern oft einer oberflächlichen Auswertung oder einer falschen Interpretation von an sich korrekten altersgruppenspezifischen Kinderlosenanteilen geschuldet sind. Insgesamt ist jedoch zu konstatieren, dass eine adäquate Einschätzung früherer Ergebnisse und exaktere Daten zu Kinderlosigkeit erst durch die MZ 2008 und 2012 möglich geworden sind.

### 3. Grenzen, Potenziale und methodische Herausforderungen des MZ

In diesem Abschnitt werden zunächst die Grenzen des MZ im Vergleich zu sozialwissenschaftlichen Datensätzen diskutiert (3.1), danach die mit der hohen Fallzahl verbundenen Potenziale (3.2) und anschließend die methodischen Aspekte von Befragung und Imputationen (3.3).

#### 3.1 *Nur Querschnittsdesign und fehlende Daten zu Männern*

Die bedeutsamste Limitierung des MZ 2012 ist, dass die Daten zur Geburtenfrage nur im Querschnitt vorliegen. Ohne ein Paneldesign sind kausale Analysen der Determinanten von Kinderlosigkeit nicht möglich – dies ist eine fundamentale Einschränkung, gerade im Vergleich zu SOEP oder pairfam. Man denke an das Zusammenspiel von Faktoren wie hohes Bildungsniveau, Wohnort in der Stadt, Partnerwahl, befristete Arbeitsverträge und Berufswechsel, die mit Kinderlosigkeit assoziiert sind. Deren Wirkung lässt sich nur in der Lebensverlaufsperspektive basierend auf Paneldaten und Ereignisanalysen identifizieren.

Eine offenkundige Einschränkung ist auch, dass Daten zu den leiblichen Kindern nur für Frauen, jedoch nicht für Männer vorliegen. Dies ist vom Gesetzgeber so gewollt (Deutscher Bundestag 2007), eine mögliche Befragung der Männer wurde beim Gesetzgebungsverfahren aus der Gesetzesvorlage gestrichen. Die Begründung, warum man Männer nicht fragen darf, ob und wie viele Kinder sie haben, ist schwer nachzuvollziehen. Die Akademiegruppe Zukunft mit Kindern empfiehlt explizit für den MZ, Fragen zu Kindern „auch männlichen Befragten“ zu stellen (Stock et al. 2012: 447).

<sup>7</sup> Hierbei sind jedoch unterschiedliche Abgrenzungen zu beachten (vgl. Scharein/Unger 2005; Schmitt/Wagner 2006; Wirth 2006; Schimpl-Neimanns 2013).

Im Jahr 1996 fand eine für den Mikrozensus im positiven Sinn gravierende Änderung statt – das traditionelle Familienkonzept wurde durch das Konzept der Lebensformen ergänzt. Damit „werden unterhalb der Ebene des Haushalts nicht Familien, sondern Lebensgemeinschaften als soziale Einheiten abgegrenzt. Diese können mit der Familie identisch sein, sind aber weiter gefasst, da sie auch unverheiratet zusammenlebende Paare einschließen.“ (Lengerer et al. 2007: 190) Allerdings ist ein Nachteil des Mikrozensus erhalten geblieben: Nach wie vor können bilokale Paarbeziehungen, Paare mit getrennten Haushalten, nicht identifiziert werden. Wer also im Mikrozensus als Single (Ein-Personen-Haushalt) oder als alleinerziehend ausgewiesen wird, muss dies de facto nicht sein.

Eine weitere Einschränkung betrifft die fehlenden Informationen über Binnenmobilität. Dadurch lassen sich u.a. Stadt-Land-Wanderungen oder Ost-West-Migration nicht erfassen, wie es beispielsweise mit einer Längsschnittstudie wie dem SOEP möglich ist. Dies sei anhand eigener Analysen verdeutlicht: Während die Kinderlosigkeit bei 41-jährigen Frauen, die zum Befragungszeitpunkt in Ostdeutschland wohnen, bei 14% liegt (nach SOEP und MZ), ist sie mit 11,5% erheblich niedriger, wenn der Wohnort vor der deutschen Wiedervereinigung berücksichtigt wird. Dies zeigt, dass nicht nur der gegenwärtige Wohnort (und dessen sozialstrukturelle und familienpolitische Rahmenbedingungen), sondern auch die kulturelle Prägung eine wichtige Rolle für das generative Verhalten spielen. Das Beispiel verdeutlicht, dass die Information über eine ostdeutsche Herkunft bei Frauen, die in Westdeutschland leben, einen analytischen Wert aufweist. Auch ist das Alter der geborenen Kinder, insbesondere das Alter der Frau bei der Geburt des ersten Kindes, beim MZ nicht direkt erfasst. Allerdings lässt es sich indirekt mit gewissen Einschränkungen erfassen, indem die Geburtenfrage mit Informationen zum Alter der Kinder im Haushalt kombiniert wird (siehe 4.3).

### *3.2 Fallzahlen und Konfidenzintervalle: Vergleich von Befunden zu Kinderlosigkeit in MZ und SOEP*

Das Potenzial des Mikrozensus für die sozialwissenschaftliche Forschung liegt insbesondere in den hohen Fallzahlen in Kombination damit, dass er auf einer Zufallsstichprobe beruht, was eine Abschätzung der Genauigkeit seiner Ergebnisse ermöglicht. Um den Einfluss unterschiedlicher Fallzahlen auf die Präzision der Prävalenzbefunde systematisch zu analysieren, werden in diesem Abschnitt die Anteile kinderloser Frauen im MZ und im SOEP für verschiedene Kohorten und sozialstrukturelle Merkmale verglichen.

Zur Analyse des Anteils kinderloser Frauen im MZ wird die Variable ef618 verwendet, die zur Frage „Haben Sie Kinder geboren?“ die Ausprägungen „Ja“, „Nein“ und „Keine Angabe“ aufweist. Eingegrenzt wird auf Frauen, die am Hauptwohnsitz und in Privathaushalten wohnen sowie für die eine Angabe zur Geburtenfrage vorliegt. Jahresüberhänge werden nicht ausgeschlossen. Das Ergebnis – nicht die Fallzahlen – wird mit dem Standardhochrechnungsfaktor auf Basis des Zensus gewichtet.

Das SOEP<sup>8</sup> ist eine repräsentative Längsschnittbefragung privater Haushalte in Deutschland, die im jährlichen Rhythmus seit 1984 durchgeführt wird (Wagner et al. 2007). Im Rahmen der SOEP-Befragung werden unter anderem die gesamten, für die Fertilitätsfor-

8 SOEP, Daten für die Jahre 1984-2012, Version 29, DOI 10.5684/soep.v29.

sung sehr zentralen Geburtsbiografien und Partnerschaftsbiografien von Frauen (und seit 2001 auch für Männer) erfasst. Da es im Verlauf einer Längsschnittbefragung zu häufigen Ausfällen kommt, wurde seit Beginn der SOEP-Befragung das Sample immer wieder durch zusätzliche Befragungsgruppen ergänzt (Kroh 2014: 8). Für den Datenabgleich mit den MZ-Daten werden in Privathaushalten lebende Frauen der Stichproben A-I verwendet. Der Vergleich der Bildungsgruppen erfolgt bei den Hochqualifizierten zwischen ISCED 5 B sowie der zusammengefassten Kategorie von ISCED 5 A und 6 (DIW/SOEP 2011: 55 ff.).

Eine zentrale Voraussetzung bei der Untersuchung von dauerhafter Kinderlosigkeit ist das Vorliegen einer vollständigen und abgeschlossenen Geburtsbiografie der Frauen. Aufgrund dessen werden alle Frauen mit unvollständigen Geburtsbiografien oder fehlenden Informationen aus der Analyse ausgeschlossen. Betrachtet werden Frauen der Geburtskohorten 1960 bis 1971, die mindestens an einer Welle in 2010 bis 2012 teilgenommen haben. Da die Geburtsbiografie der Frauen im SOEP jährlich aktualisiert wird und somit nur vollständig sein kann, wenn die Befragte im Alter von 41 Jahren, also am Ende der hier betrachteten Periode, an der Befragung teilgenommen hat, werden auch nur diese Frauen in der Analyse berücksichtigt (vgl. Frick/Schmitt 2013: 90).<sup>9</sup> Durch die hier aufgeführten Eingrenzungen werden 723 Fälle ausgeschlossen und es verbleiben 3.185 Frauen im Analysesample.

Tabelle 1 zeigt die Kinderlosigkeit im Vergleich von SOEP und MZ nach Kohorten, Bildung, Region und Migrationshintergrund. Bei den hier verglichenen Frauen der Jahrgänge 1960-71 kann weitestgehend davon ausgegangen werden, dass sie ihre fertile Phase abgeschlossen haben.

*Tabelle 1:* Kinderlosigkeit von Frauen im Vergleich von SOEP und MZ

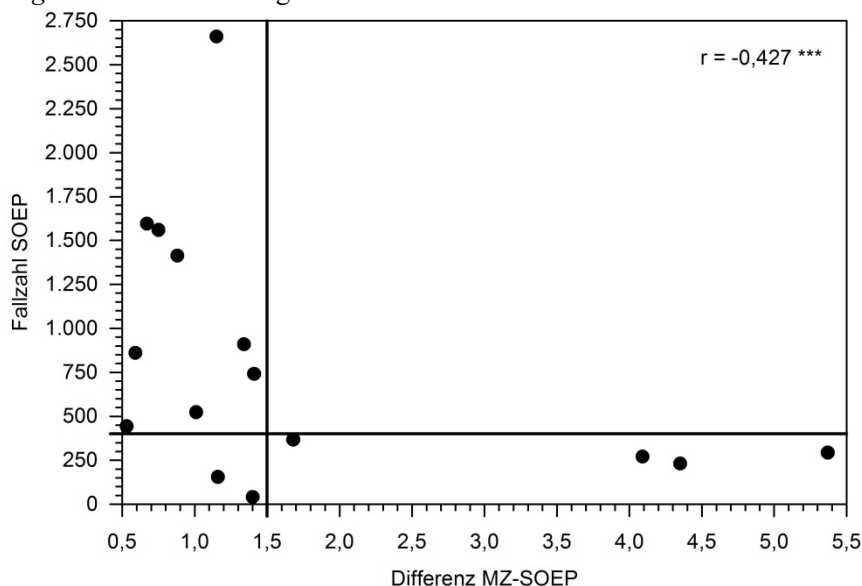
	SOEP		MZ 2012		Differenz (in Prozent- punkten)
	Anteil kinder- los (in %)	N	Anteil kinder- los (in %)	N	
1960 bis 1962	17,34	861	17,93	16.223	0,59
1963 bis 1967	20,47	1.414	19,59	28.001	-0,88
1968 bis 1971	22,57	910	21,23	19.437	-1,34
ISCED 1 (ohne Abschluss)	14,56	42	15,96	2.019	1,40
ISCED 2 (Haupt-/Realschulabschluss)	10,76	271	14,85	6.717	4,09
ISCED 3 (Lehre oder Hochschulreife)	18,33	1.596	17,66	33.165	-0,67
ISCED 4 (Lehre und Hochschulreife)	20,25	294	25,62	5.657	5,37
ISCED 5 B (Meister)	25,66	232	21,31	7.053	-4,35
ISCED 5 A u. 6 (Hochschulabschluss)	28,17	742	26,76	8.982	-1,41
Westdeutschland (Wohnort mit 41 J.)	21,77	1.560	21,02	51.111	-0,75
Ostdeutschland (Wohnort mit 41 J.)	13,60	444	14,13	12.550	0,53
Westdeutschland (Wohnort 1989)	23,80	2.240	–	–	–
Ostdeutschland (Wohnort 1989)	11,50	741	–	–	–
ohne Migrationshintergrund	22,42	2.661	21,27	53.563	-1,15
mit Migrationshintergrund	12,87	524	11,86	10.098	-1,01
1. Migrantengeneration	9,51	368	11,19	9.617	1,68
2. Migrantengeneration	23,73	156	24,89	481	1,16

*Quelle:* MZ 2012 und SOEP, eigene Berechnung.

9 Es wurden verschiedene Altersgrenzen (39 bis 45 Jahre) getestet. Es zeigen sich keine gravierenden Unterschiede bei der prozentualen Verteilung der Kinderlosigkeit im Analysesample: Bei Rechtszensurierung von 41 Jahren liegt die Kinderlosigkeit bei 20,35%, bei 42 Jahren 20,29% und bei 43 Jahren 20,06%.

Die Differenzen der Kinderlosigkeit zwischen SOEP und MZ sind teilweise gering, teilweise aber sehr hoch. Sie liegen zwischen 0,53 und 5,37 Prozentpunkten. Die hohen Differenzen liegen ausnahmslos in den Gruppen, in denen die Fallzahl in einem der Datensätze (hier SOEP) gering ist. Die Hilfslinien in Abbildung 3 verdeutlichen, dass bei  $N > 400$  die Differenz immer unter 1,5 Prozentpunkten liegt und bei geringeren Fallzahlen teilweise deutlich darüber. Zweifellos gibt es auch andere erhebungsspezifische Gründe und Selektionseffekte, die zu Unterschieden in der Kinderlosigkeit zwischen SOEP und MZ führen. Hier wäre exemplarisch die unterschiedliche Erfassung der Bildung in beiden Datensätzen zu nennen. Schimpl-Neimanns (2013) hat in seinen Untersuchungen gezeigt, dass es mit MZ zu einer Überschätzung der Real- und Gymnasialschüler bei einer gleichzeitigen Unterschätzung von Schülern in beruflichen Gymnasien kommt (ibid.: 53). Der Hauptunterschied ist jedoch auf geringe Fallzahlen zurückzuführen. Ihre Korrelation mit der Differenz ist hochsignifikant ( $r = -0,427$ ).

Abbildung 3: Zusammenhang zwischen Fallzahl und SOEP-MZ-Differenzen



Quelle: Datenbasis MZ 2012 und SOEP, eigene Berechnung und Darstellung.

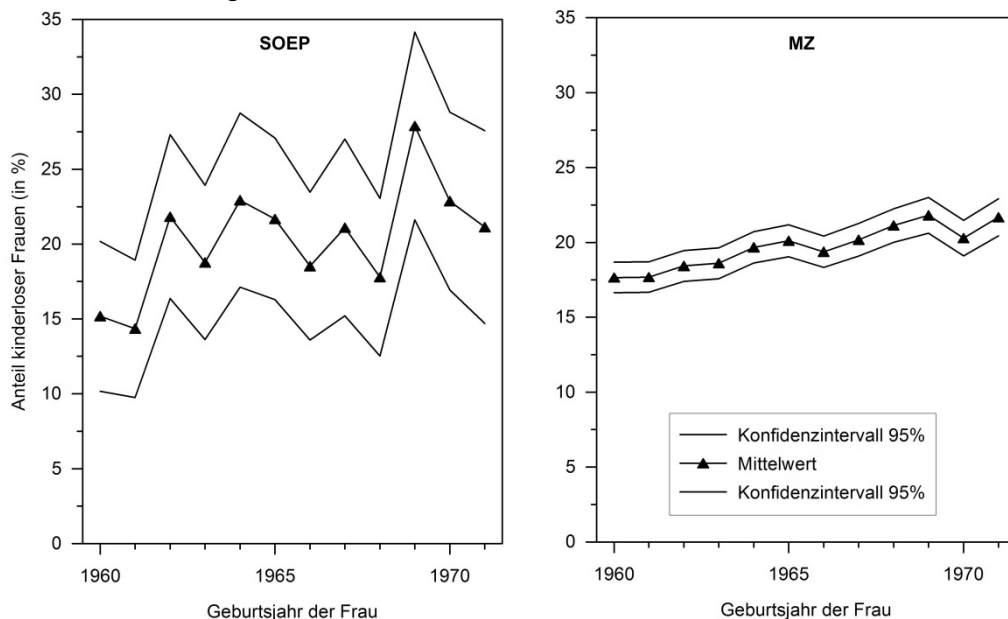
Um den Einfluss von Fallzahlen auf die Verlässlichkeit von Fertilitätsmerkmalen, v.a. soziostrukturell differenzierten, systematisch zu erfassen, werden die Konfidenzintervalle für die Kinderlosigkeit von Frauen der Kohorten 1960-1971 berechnet (vgl. Häder/Häder 2014; Kühnel/Krebs 2010).<sup>10</sup> Während die Konfidenzintervalle beim SOEP mit der STATA-Prozedur *ci* und einer binären Kodierung berechnet wurden, wurde beim MZ die SAS-Prozedur „*surveymeans*“ verwendet (SAS Institute 2008). Die Prozedur ermöglicht bei Varianzberechnungen die statistischen Eigenschaften komplexer Stichproben (im MZ

10 Zum statistischen Zusammenhang von Fallzahl und Konfidenzintervall siehe u.a. Häder und Häder (2014). Für die statistische Berechnung der Konfidenzintervalle siehe u.a. Kühnel und Krebs (2010).

liegt eine einfach geschichtete Klumpenstichprobe vor) sowie die Endlichkeitsfaktoren von Stichproben aus finiten Populationen zu berücksichtigen (Schimpl-Neimanns 2011). Die Fallzahlen beim SOEP betragen im Durchschnitt der ausgewählten Frauenkohorten ungewichtet 265 und beim MZ 5.324. Es wurden gewichtete Werte berechnet, die jedoch fallzahlkorrigiert sind. Der Standardfehler des Mittelwerts liegt beim SOEP bei 0,0280 und beim MZ bei 0,0056.

Die Verläufe der Kinderlosigkeit von Frauen im Kohortenvergleich von 1960 bis 1971 zeigen einen weitaus gleichmäßigeren Verlauf beim MZ und deutliche Schwankungen beim SOEP. Noch deutlicher zeigt sich der Unterschied, wenn man die Konfidenzintervalle betrachtet (Abb. 4): Der Abstand der Intervallgrenzen zum Mittelwert liegen beim SOEP zwischen 4,6 und 6,4 und beim MZ zwischen 1,0 bis 1,2 Prozentpunkten (vgl. Schupp et al. 1997). Auch die in Abb. 3 gezeigten Differenzen zwischen SOEP und MZ liegen innerhalb des Konfidenzintervalls beim SOEP. Während bei Fallzahlen von 200 eher eine Zusammenlegung von Ausprägungen (bspw. Fünf-Jahres-Kohorten) angebracht ist, ermöglicht der MZ noch über den einzelnen Jahrgang hinausgehende Differenzierungen.

Abbildung 4: Konfidenzintervalle zum Anteil kinderloser Frauen 1960-1971 im Vergleich von SOEP und MZ



Quelle: Datenbasis MZ 2012 und SOEP, eigene Berechnung und Darstellung.

Angesichts dieses Vergleichs von MZ und SOEP ist zu betonen, dass beide Datensätze unterschiedliche Zielsetzungen haben. Ziel des SOEP ist es, eine Datengrundlage für sozial- und wirtschaftswissenschaftliche Analysen, insbesondere lebenslaufbezogenem Verhaltens zu schaffen (Krupp 2008). Um derartige Zusammenhänge zu analysieren, sind kleine Konfidenzintervalle bezüglich der Punktprevalenz nicht notwendig, sondern ein Längsschnittdesign (siehe 3.1). Der Mikrozensus dagegen hat das Ziel, in kurzen und re-

gelmäßigen Abständen detaillierte Daten über die Bevölkerungsstruktur, die wirtschaftliche und soziale Lage der Bevölkerung etc. bereitzustellen und dadurch die Lücke zwischen zwei Volkszählungen zu füllen (Statistisches Bundesamt 2013b). Hierbei entstehen Analysepotenziale für tiefenscharfe Differenzierungen von Sozialstruktur und Zeitverläufen.

### 3.3 Methodische Aspekte bezüglich Erhebung und Imputationen beim MZ

In diesem Abschnitt werden die Art der Beteiligung an der Erhebung, Antwortverweigerer, Imputation und Jahresüberhänge in Hinblick auf potenzielle Verzerrungen diskutiert (zu stichprobenbedingten Fehlerquellen und Methodik siehe: Statistisches Bundesamt 2013b).

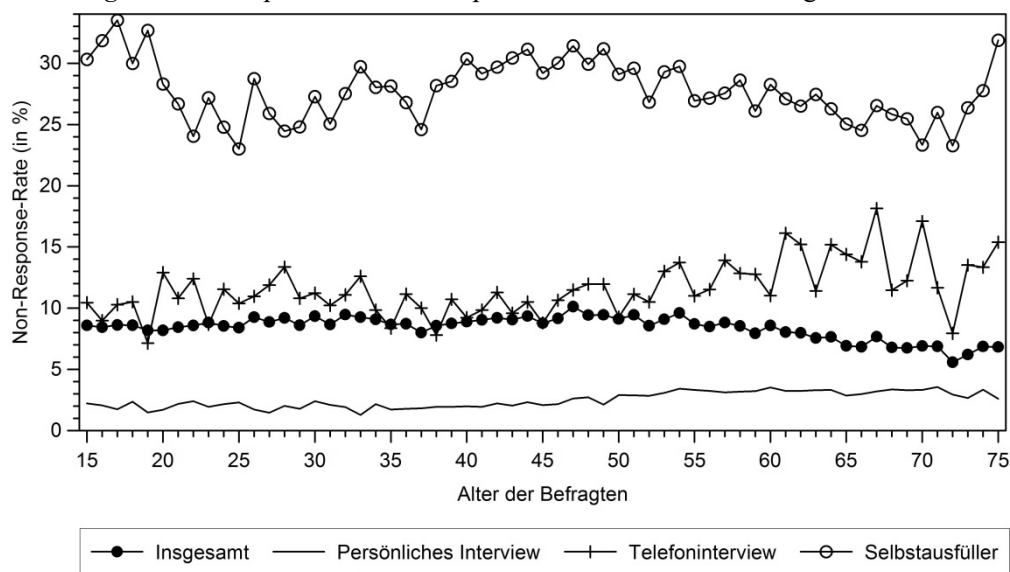
#### 3.3.1 Art der Beteiligung an der Erhebung und unterschiedliche Antwortverweigerungen

Die Art der Erhebung hat einen bedeutsamen Einfluss auf die Beantwortung der Frage zur Geburt von Kindern. Sie lässt sich zweifach differenzieren:

- Ob durch persönliches Interview, Telefoninterview oder Selbstausfüller-Haushalt und
- ob selbst beantwortet, Proxyangabe oder keine Antwort.

Nicht beantwortet wurde die Frage nach geborenen Kindern überwiegend von Selbstausfüller-Haushalten. Bei diesen lag die altersspezifische Non-Response-Rate im MZ 2012 zwischen 23,0% und 32,7%. Im persönlichen Interview variiert dieser Wert zwischen 1,3% und 3,6%. Abbildung 5 verdeutlicht diese großen Unterschiede und auch, dass mit zunehmendem Alter die Antwortverweigerer insgesamt abnehmen. Eine erhöhte Antwortverweigerung am Ende der reproduktiven Phase ist nicht gegeben.

Abbildung 5: Altersspezifische Non-Response-Rate nach der Erhebung im MZ 2012



Quelle: Datenbasis MZ 2012, eigene Berechnung und Darstellung.

Nach Möglichkeit soll jede Person im Alter von 15 Jahren oder älter die Fragen des MZ selbst, gegebenenfalls mit Unterstützung einer weiteren Person, beantworten. Es sind aber auch so genannte Proxy-Interviews zulässig, das heißt ein Haushaltsmitglied darf stellvertretend für andere Haushaltsmitglieder antworten (zur Durchführung siehe: Statistische Ämter des Bundes und der Länder, o.J.). Fremdauskünfte liegen im MZ 2012 für etwa 26% der zu befragenden Personen im Alter von 15 Jahren oder älter vor.

Werden die Angaben zur Geburt eines Kindes nach der Art der Beteiligung an der Erhebung differenziert, zeigen sich erhebliche Unterschiede: Während die Kinderlosigkeit bei Selbstauskunft 26,1% beträgt, sind es bei Proxyangabe 42,7%. Personen, die eine freiwillige Frage nicht beantworten, machen auch bei anderen freiwilligen Fragen von ihrem Verweigerungsrecht Gebrauch. Von den Befragten, für die keine Angaben zur Art der Beteiligung an der Erhebung vorliegen, verweigerten 49,5% auch die Antwort zur Geburt von Kindern.

Welche Verzerrungen könnten diese sehr unterschiedlichen Non-Response-Raten mit unterschiedlichen Kinderlosigkeitsraten verursachen? Bei fehlenden Werten könnte dann eine Unterschätzung vorliegen, wenn kinderlose Frauen die Frage (beispielsweise aus Scham) eher verweigerten (Kreyenfeld et al. 2011: 389). Dann wäre aber auch zu erwarten, dass im Alter um 40 Jahre, wenn sich temporäre Kinderlosigkeit im Lebensverlauf als endgültige herausstellt, die Antwortverweigerung ansteigt, was nicht der Fall ist (Abb. 5). Bei den Proxyangaben wird die Kinderlosigkeit tendenziell überschätzt, da der Auskunftgebende in einigen Fällen nicht von vorhandenen Kindern weiß. Da Proxyangaben in einem engen Zusammenhang mit Merkmalen der Sozialstruktur stehen (Wolf/Lüttinger 2009), ist auch eine Überschätzung in verschiedenen Bevölkerungsteilen gegeben. Bemerkenswert sind die Angaben hinsichtlich kinderloser Frauen, wenn also die Frage zur Geburt von Kindern verneint wurde. Wird nur zwischen der Art der Beteiligung unterschieden, geben 26,1% der Frauen, die selbst Auskunft geben, an, kinderlos zu sein. Liegt eine Fremdauskunft vor, steigt dieser Anteilwert auf 42,7% an. Diese Unterschiede könnten zu systematischen Verzerrungen führen. Erklärungen für den Zusammenhang zwischen Proxyangabe und höherer Kinderlosigkeit könnten die Hintergrundfaktoren Erwerbstätigkeit und Lebensform sein:

- (Vollzeit) erwerbstätige Frauen haben eine höhere Kinderlosigkeit und sind seltener zur Selbstauskunft bereit. Analoges gilt für Selbstausfüller: Dies sind überproportional erwerbstätige Frauen, die für ein persönliches Interview nicht anzutreffen sind.
- Wie genau ein Partner von der Mutterschaft weiß, hängt von der Lebensform ab. Werden nur Paargemeinschaften betrachtet, ist bei verheiratet zusammenlebenden Ehepaaren kein bedeutsamer Unterschied zwischen Selbst- und Fremdauskunft festzustellen (11,2% und 11,6%). Bei allen anderen Formen einer Partnerschaft geben 45,8% der Frauen an, kinderlos zu sein. Gibt der Partner Auskunft, liegt der Anteil kinderloser Frauen bei 54,1%.

Je stärker hier die fehlenden Angaben bei der Imputation reduziert werden können, desto geringer ist der Verzerrungseffekt einer Unterschätzung von Kinderlosigkeit.

### 3.3.2 Antwortverweigerungen und Imputationen

Da die Beantwortung des MZ verpflichtend ist, stellen Verzerrungen durch Antwortverweigerer kein so großes Problem dar, wie es bei freiwilligen Surveys der Fall ist. Die Unit-Non-Response liegt bei 2,1% (Statistisches Bundesamt 2013b: 7); dies ist ein immenser Vorteil des MZ. Anders sieht es bei der Beantwortung von freiwilligen Fragen aus, wie bei der Frage nach dem „ob“ und der Zahl der Kinder.

Der Anteil der Antwortverweigerer zur Geburtenfrage im MZ 2008 betrug 9,6%<sup>11</sup> und 2,2% der zu befragenden Frauen wurde die Frage nicht gestellt, da sie mit dem Fragebogen aus 2007 befragt wurden. In 2012 fällt das Niveau der Antwortverweigerer mit 8,4% geringer aus. Nach den Imputationen reduziert sich der Anteil der Antwortverweigerer im Jahr 2012 von 8,4% auf 3,7%. Die Antwortausfälle können sich strukturverzerrend auswirken, wenn es Hinweise auf systematische Ausfälle gibt. Vergleiche zwischen Frauen im Alter von 15 und 75 Jahren, die keine Angaben zu geborenen Kindern machten, die durch Interviewer befragt wurden bzw. den Fragebogen schriftlich selbst ausfüllten, zeigen auf, dass Frauen in Selbstausfüller-Haushalten

- häufiger ledig sind: Anteil lediger Frauen im persönlichen Interview: 26,2%, im Selbstausfüller-Haushalt: 35,5%;
- über ein höheres Bildungsniveau verfügen: Anteil der zusammengefassten ISCED-97-Stufen 5b, 5a, und 6 im persönlichen Interview: 18,4%, im Selbstausfüller-Haushalt: 28,8% und
- jünger sind: Durchschnittsalter von Frauen im Alter von 15 bis 75 Jahren im persönlichen Interview: 47,6 Jahre, im Selbstausfüller-Haushalt: 43,6 Jahre.

Aufgrund des Vergleichs dieser wenigen Strukturmerkmale wurde seitens des Statistischen Bundesamtes (2009b: 2) angenommen, dass „Frauen, die nicht antworten, (...) mit hoher Wahrscheinlichkeit überdurchschnittlich häufig kinderlos“ sind. Dies entspricht den Überlegungen in 3.3.1.

Zur Korrektur des Non-Response wurde ein zweistufiges Imputationsverfahren eingesetzt. Imputationen sind bei Frauen sinnvoll, bei denen aus vorliegenden Informationen mit einer über 90-prozentigen Wahrscheinlichkeit angenommen werden, dass sie kinderlos sind bzw. sich die Kinderzahl bestimmen lässt. In der ersten Stufe wird die Geburt von Kindern geschätzt und in der zweiten Stufe die Zahl der geborenen Kinder (Statistisches Bundesamt 2009b).

In der ersten Imputationsstufe gelten Frauen mit Kindern im Haushalt als Mütter. Zwar wird im MZ bei den Fragen, ob die Mutter bzw. der Vater im Haushalt lebt, nicht zwischen leiblichen, Pflege-, Stief- und Adoptivkindern als Kinder einer Frau eines Mannes unterschieden, sodass Frauen, die selbst kein Kind geboren haben, fälschlicherweise eine Mutterschaft unterstellt wird. Da aber nur bei 0,6% der Frauen, die angaben kein Kind geboren zu haben, Kinder leben, wurde der Fehler toleriert. Für Frauen ohne im

---

11 Im Gegensatz zum Methodenbericht zur Imputation von Werten bei fehlenden Angaben zur Mutterschaft (Statistisches Bundesamt 2009b) werden hier nicht die Anteilwerte insgesamt genannt, sondern die Angaben beziehen sich nur auf Frauen, die mit dem Frageprogramm des MZ 2008 bzw. des MZ 2012 befragt wurden. Dies gilt auch für alle weiteren Angaben in diesem Abschnitt.



Haushalt lebende Kinder wurde die Wahrscheinlichkeit einer Kinderlosigkeit unter Berücksichtigung der Merkmale Alter, Familienstand und Bildungsniveau geschätzt.

In der zweiten Stufe der Imputation wird die Zahl der geborenen Kinder geschätzt. Auch hier wurde ein Wahrscheinlichkeitsmodell entworfen, das auf den Daten der Befragten mit Angaben zu Zahl geborener Kinder beruht. Ausgangsbasis sind Frauen, deren Zahl geborener Kinder mit der Zahl der im Haushalt lebenden Kinder übereinstimmt. Aus der Kombination von Bildungsniveau, Gebietsstand und Altersgruppen der Frauen wurde dann die Zahl der im Haushalt lebenden Kinder der Frau als Zahl geborener Kinder zugewiesen. Bei der Kinderzahl ist die Imputation jedoch weniger effektiv als bei der Frage, ob eine Frau Kinder hat. Auch weil einige Frauen die Frage des „ob“, jedoch nicht die nach dem „wie viel“ beantwortet haben, sind die Werte zur Kinderlosigkeit exakter als die zur durchschnittlichen Kinderzahl.

Tabelle 2 gibt einen Überblick zu den Anteilwerten, differenziert nach Art der Beteiligung an der Erhebung und Art der Erhebung zur Frage der geborenen Kinder vor und nach der Imputation. Eine nennenswerte Abweichung nach der Imputation zwischen den Frauen insgesamt und den Frauen, die nicht dem Jahresüberhang angehören, ist nur bei den Selbstausfüller-Haushalten gegeben – die noch mit dem Fragebogen 2011 befragt wurden.

**Tabelle 2:** Angaben zur Geburt von Kindern vor und nach Imputation, differenziert nach Art der Beteiligung an der Erhebung und Art der Erhebung

Art der Beteiligung an der Erhebung	N	Ja	Nein	Kind(er) geboren			Ja	Nein	k. A.
-----				k. A.	N				
Art der Erhebung	Fallzahl			Prozent					
Vor Imputation: EF60 (mit Fragebogen aus 2012 befragt)									
Selbstauskunft	184.859	127.312	48.218	9.329	100	68,9	26,1	5,0	
Proxyangabe	59.156	30.277	25.248	3.631	100	51,2	42,7	6,1	
Keine Angabe	18.598	5.416	3.978	9.204	100	29,1	21,4	49,5	
Persönliches Interview	186.921	127.344	54.738	4.839	100	68,1	29,3	2,6	
Telefoninterview	23.349	13.093	7.569	2.687	100	56,1	32,4	11,5	
Selbstausfüller-Haushalt	52.343	22.568	15.137	14.638	100	43,1	28,9	28,0	
Insgesamt	262.613	163.005	77.444	22.164	100	62,1	29,5	8,4	
Nach Imputation: EF618 (insgesamt)									
Selbstauskunft	187.341	130.377	51.048	5.916	100	69,6	27,3	3,2	
Proxyangabe	59.778	31.495	26.597	1.686	100	52,7	44,5	2,8	
Keine Angabe	19.936	9.446	7.143	3.347	100	47,4	35,8	16,8	
Persönliches Interview	186.921	128.344	55.719	2.858	100	68,7	29,8	1,5	
Telefoninterview	23.435	13.853	8.283	1.299	100	59,1	35,3	5,5	
Selbstausfüller-Haushalt	56.699	29.121	20.786	6.792	100	51,4	36,7	12,0	
Insgesamt	267.055	171.318	84.788	10.949	100	64,2	31,8	4,1	

*Quelle:* MZ 2012, eigene Berechnung.

*Anmerkungen:* k.A. = keine Angabe. Die Angaben zu der Frage zur Geburt von Kindern vor Imputation bezieht sich auf Frauen im Alter von 15 bis 75 Jahren, die mit dem Frageprogramm des MZ 2012 befragt wurden, die Angaben nach Imputation auf die Frauen im Alter von 15 bis 75 Jahren insgesamt.

Durch die Imputationen werden die Daten zwar nicht inhaltlich verändert, da sie auf Nebenbedingungen beruhen, die durch die anderen Ergebnisse gespeist sind. Grundsätzlich erhöht sich jedoch die Fallzahl und es reduziert sich die Black Box der Ungewissheit über Befragte, deren Kinderzahl unbekannt ist. Dadurch ist die potenzielle Verzerrung geringer. Wichtiger ist jedoch, dass die Unterschätzung der Kinderlosigkeit korrigiert wird, die unserer Hypothese nach mit der Erwerbstätigkeit der Frauen verbunden ist. Dies betrifft bei der Beteiligung die Proxyangabe und bei Selbstausfüller-Haushalten: Tabelle 2 zeigt, wie deren fehlenden Werte zur Frage, ob Kinder geboren sind, überproportional durch die Imputation reduziert wurden. Entsprechend hat sich die Kinderlosigkeit aller Frauenjahrgänge zusammen von 29,5% auf 31,8% erhöht. Nach der Imputation liegt die Kinderlosigkeit bei Proxyangabe mit 44,5% und bei Selbstausfüllerhaushalten mit 36,7% immer noch deutlich über dem Durchschnitt. Allerdings sind die fehlenden Werte bei Proxyangaben sogar unter der Selbstauskunft und auch bei Selbstausfüllern von 28% auf 12% reduziert. Folglich konnte ein erheblicher Teil der Verzerrung durch die Imputation reduziert werden, wobei eine geringe tendenzielle Unterschätzung der Kinderlosigkeit noch bleibt.

### 3.3.3 Verbesserungen der Befragung beim MZ 2012 im Vergleich zum MZ 2008

Aufgrund des relativ hohen Anteil von 9,6% Antwortverweigerinnen zur Frage nach geborenen Kindern<sup>12</sup> und der mehrheitlich geäußerten Kritik über die Platzierung der Fragen zu den geborenen Kindern am Ende des Fragebogens im MZ 2008 (Kreyenfeld/Konietzka 2013, Stock et al. 2012: 447), wurden die Fragen im MZ 2012 in den Familienkontext zu Beginn des Fragebogens eingebettet. In der Gesamtbetrachtung ist zwar eine Verbesserung der Non-Response-Rate um 1,2 Prozentpunkte festzustellen, die Verbesserung fiel aber geringer aus als erhofft (Tab. 3).

*Tabelle 3:* Non-Response-Rate zur Frage nach geborenen Kindern in MZ 2008 und 2012

Art der Erhebung	Non-Response-Rate (in %)	
	MZ 2008	MZ 2012
Persönliches Interview	1,3	2,6
Telefoninterview	15,6	11,5
Selbstausfüller-Haushalte	43,1	28,0
Insgesamt	9,6	8,4

*Quelle:* MZ 2008 und MZ 2012, eigene Berechnung.

Allerdings führte in den Bereichen der Telefoninterviews und insbesondere der Selbstausfüller-Haushalte die Einbettung der Fragen in den Familienkontext zu einer deutlichen Verbesserung in der Beantwortung der Frage zu den geborenen Kindern. Die Non-Response-Rate reduzierte sich bei den Telefoninterviews um 4,1 Prozentpunkte und bei den Selbstausfüller-Haushalten sogar um 15,1 Prozentpunkte. Dadurch hat sich die potenzielle Verzerrung durch die höhere Kinderlosigkeit dieser Gruppen reduziert.

12 Die Non-Response-Rate bezieht sich auf die Frauen im Alter von 15 bis 75 Jahren, die mit dem Frageprogramm des MZ 2012 befragt wurden. Auch alle weiteren Non-Response-Raten in diesem Abschnitt beziehen sich auf die Frauen im Alter von 15 bis 75 Jahren, die mit dem Frageprogramm des MZ 2012 befragt wurden.

Insgesamt ist die Verzerrung durch Antwortverweigerer als relativ gering einzuschätzen. Denn durch die Antwortpflicht des MZ sind Verzerrungen bei der Generierung von befragten Personen sehr gering – ein großer Vorteil gegenüber sozialwissenschaftlichen Erhebungen. Auch ist die Zahl der Antwortverweigerer auf diese freiwillige Frage im MZ 2012 gesunken, und durch Imputationen kann der Anteil der Antwortverweigerer auf 3,7% reduziert werden.

### 3.3.4 Problematik der Jahresüberhänge

Jahresüberhänge stellen dann ein Problem dar, wenn sie noch mit dem Fragebogen aus dem Vorjahr befragt wurden. In den Publikationen des Statistischen Bundesamtes (u.a. 2009a, 2009b, 2013a) werden Jahresüberhänge kaum thematisiert. Bis zum Jahr 2004 wurde der MZ mit einer festen Berichtswoche erhoben, die in der Regel in der letzten feiertagsfreien Woche im April lag. Seit 2005 erfolgte eine Umstellung auf ein unterjähriges, kontinuierliches Erhebungsverfahren, das eine gleichmäßige Verteilung der Berichtswochen über das gesamte Jahr beinhaltet. Die Befragung auskunftspflichtiger Haushalte kann aber nicht stets an den, gemäß Auswahlplan, festgelegten Befragungsterminen eingehalten werden, sodass es zu zeitlichen Verschiebungen in einen späteren Referenzzeitraum kommt. Liegt dieser im Folgejahr und die Antworten werden im Folgejahr berücksichtigt, liegt ein so genannter Jahresüberhang vor. Solche Verschiebungen wirken sich unmittelbar auf die Fragen der im vierjährigen Turnus wechselnden Zusatzprogramme – wie der Geburtenfrage – aus, wenn die Auskunft aus einem Selbstausfüller-Haushalt vorliegt. Wird der Fragebogen im Dezember eines Jahres versandt und der auskunftspflichtige Haushalt sendet diesen Mitte Januar oder später an das zuständige Statistische Landesamt zurück, werden die Antworten im aktuellen Jahr berücksichtigt. Erhebungsbedingt liegen für diese Selbstausfüller-Haushalte keine Angaben zu dem im aktuellen Jahr eingesetzten Zusatzprogramm vor. Im Datenfile des MZ werden solche Fälle als Item-Non-Response gekennzeichnet. Analysen zu Antwortverweigerungen müssen diesen Aspekt stets berücksichtigen, da Personen keine Antworten verweigern können, wenn sie die Frage nicht gestellt bekommen haben. Im Methodenbericht zur Imputation von Werten (Statistisches Bundesamt 2009b) wird beispielsweise diese Differenzierung nicht vorgenommen und von einer Non-Response-Rate von rund 12% berichtet. Wird die Non-Response-Rate nur auf die Frauen bezogen, die mit dem Frageprogramm 2012 befragt wurden, reduziert sich die Zahl der Antwortverweigerer um knapp 2 Prozentpunkte: 2008 von 11,57 auf 9,71% und 2012 von 10,08 auf 8,44%.

Seitens der Wissenschaft werden die Jahresüberhänge zu den Fragen, ob und wie viele Kinder geboren wurden, generell mit einbezogen. Dies ist, bezogen auf die Geburt von Kindern, unproblematisch, da Jahresüberhänge in den Daten des MZ als fehlende Angaben kodiert werden und diese im Nachgang soweit möglich imputiert werden.

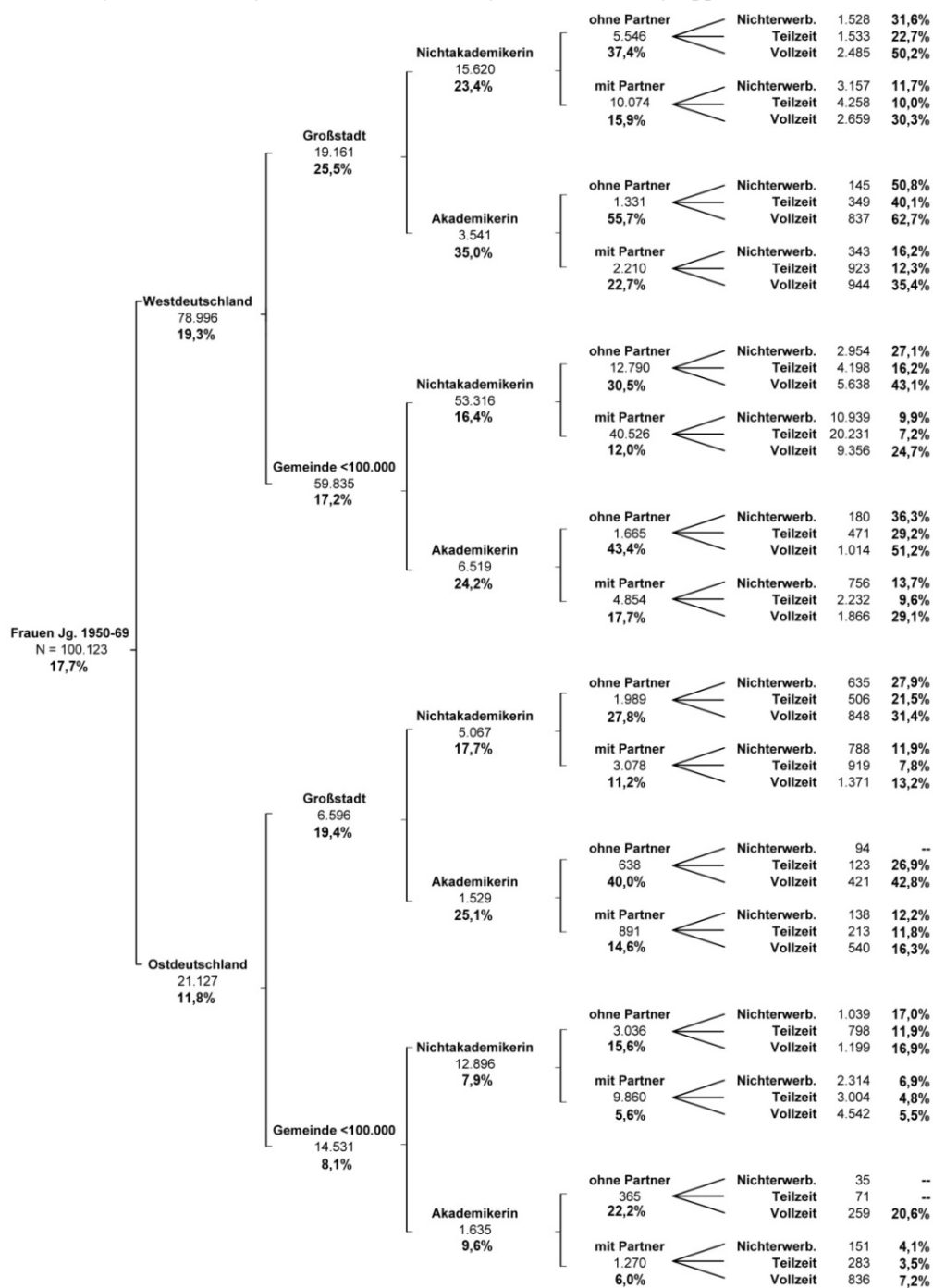
#### **4. Drei Beispiele einer differenzierten Erfassung von Sozialstruktur, Paritäten und Zeitverläufen**

Das Potenzial der hohen Fallzahlen des MZ macht eine Differenzierung nach verschiedenen soziodemografischen und sozioökonomischen Gruppen, nach Paritäten und nach einzelnen Jahrgängen möglich. So lassen sich detaillierte Entwicklungen und Strukturen identifizieren. Durch die Verfügbarkeit von Daten seit 1973 ist es möglich, Strukturveränderungen über mehr als vier Jahrzehnte zu analysieren und Trenddatenanalysen durchzuführen (vgl. Schimpl-Neimanns 2011: 37). Lengerer et al. (2007: 187) verweisen darauf, dass der Mikrozensus eine Mehrzweckstichprobe ist, die zwar nicht spezifisch für familienwissenschaftliche Fragestellungen entwickelt wurde, aber dennoch ein beträchtliches Potenzial für familiensoziologische und demografische Analysen besitzt. Diese Potenziale können anhand von deskriptiven statistischen Methoden, aber auch anhand von strukturprüfenden Verfahren wie linearen oder logistischen Regressionen gehoben werden – Voraussetzung ist ein entsprechendes Forschungsdesign. Im Folgenden werden Beispiele dafür anhand der drei vorne genannten Dimensionen gezeigt: Kinderlosigkeit nach sozialstruktureller Differenzierung anhand einer Dendrogrammanalyse (4.1), paritätsspezifische Differenzierung in Kombination mit Bildungsgruppen (4.2) und kohortenspezifische Entwicklung des Erstgebäralters nach Bildungsgruppen (4.3).

##### *4.1 Sozialstrukturelle Differenzierung: Kinderlosigkeit nach verschiedenen Merkmalskombinationen*

Im Folgenden werden die Möglichkeiten eines Dendrogrammes am Beispiel der Kinderlosigkeit für verschiedene Merkmalskombinationen dargestellt. Die Differenzierung bis zur fünften Ebene ermöglicht für die Kombination von vier dichotomen Gruppen West-/Ostdeutschland, Großstadt/Gemeinde mit unter 100.000 Einwohnern, Nicht-/Akademikerinnen, mit/ohne Partner im Haushalt sowie der Differenzierung zwischen Nicht-, Teilzeit- und Vollzeit-erwerbstätigen 48 verschiedene Subgruppen. Die Reihenfolge der Aufteilung berücksichtigt, dass die ersten Merkmale stabiler sind und eher Hinweise auf die Ursachen von Kinderlosigkeit geben, während Partnerschaftsstatus und Arbeitszeit auch als Folgen von Mutterschaft bzw. endgültiger Kinderlosigkeit interpretiert werden können. Abbildung 6 zeigt ein Dendrogramm, das die subgruppenspezifische Kinderlosigkeit und die Fallzahlen zeigt.

Abbildung 6: Dendrogramm von Kinderlosigkeit nach Untergruppen



Quelle: Eigene Berechnung und Darstellung basierend auf MZ 2012.

*Anmerkungen:* Unterhalb der Kategorien stehen die Fallzahlen, darunter die gewichtete Kinderlosenquote. Der Gewichtungsfaktor basiert auf der Fortschreibung der Bevölkerungsstatistik (Stand Juni 2015). „Nichterwerb.“ umfasst die nichterwerbstätigen und arbeitslosen Frauen. „Ohne Partner“ bedeutet ohne Partner im Haushalt. Bei  $N < 100$  sind keine Werte angegeben.

Die Kinderlosenquote von 17,7% aller Frauen der Jahrgänge 1950-69 teilt sich auf 19,3% in West- und 11,8% in Ostdeutschland auf. Wie sieht die Verästelung bei Gruppen mit niedriger Kinderlosigkeit aus? Der Anteil von 11,8% in Ostdeutschland variiert weniger nach Bildung, aber erheblich im regionalen Vergleich: In Großstädten ist sie mit 19,4% mehr als doppelt so hoch wie in kleineren Städten oder ländlichen Kreisen mit 8,1%. Mit Partner liegt die Kinderlosigkeit bei ostdeutschen Nichtakademikerinnen in weniger urbanen Gemeinden bei 5,6%. Ohne Partner liegt sie bei Akademikerinnen in Großstädten bei 55,7%, wenn sie Vollzeit arbeiten sogar bei 62,7%. Im Unterschied zu Westdeutschland sind die Unterschiede nach der Arbeitszeit in Ostdeutschland gering, bei Teilzeit liegt die Kinderlosigkeit bei 4,8%, bei Vollzeit 5,5%. Am niedrigsten ist die Kinderlosigkeit mit 3,5% bei ostdeutschen Akademikerinnen, die mit Partner in weniger urbanen Gemeinden leben und in Teilzeit arbeiten. Die niedrigste Kinderlosigkeit in Westdeutschland mit 7,2% haben Nichtakademikerinnen in weniger urbanen Gemeinden mit Partner, die in Teilzeit arbeiten. Bemerkenswert ist, dass dies die mit Abstand größte von allen 48 Gruppen der fünften Subebene ist, sie macht mehr als 20% aller Frauen aus.

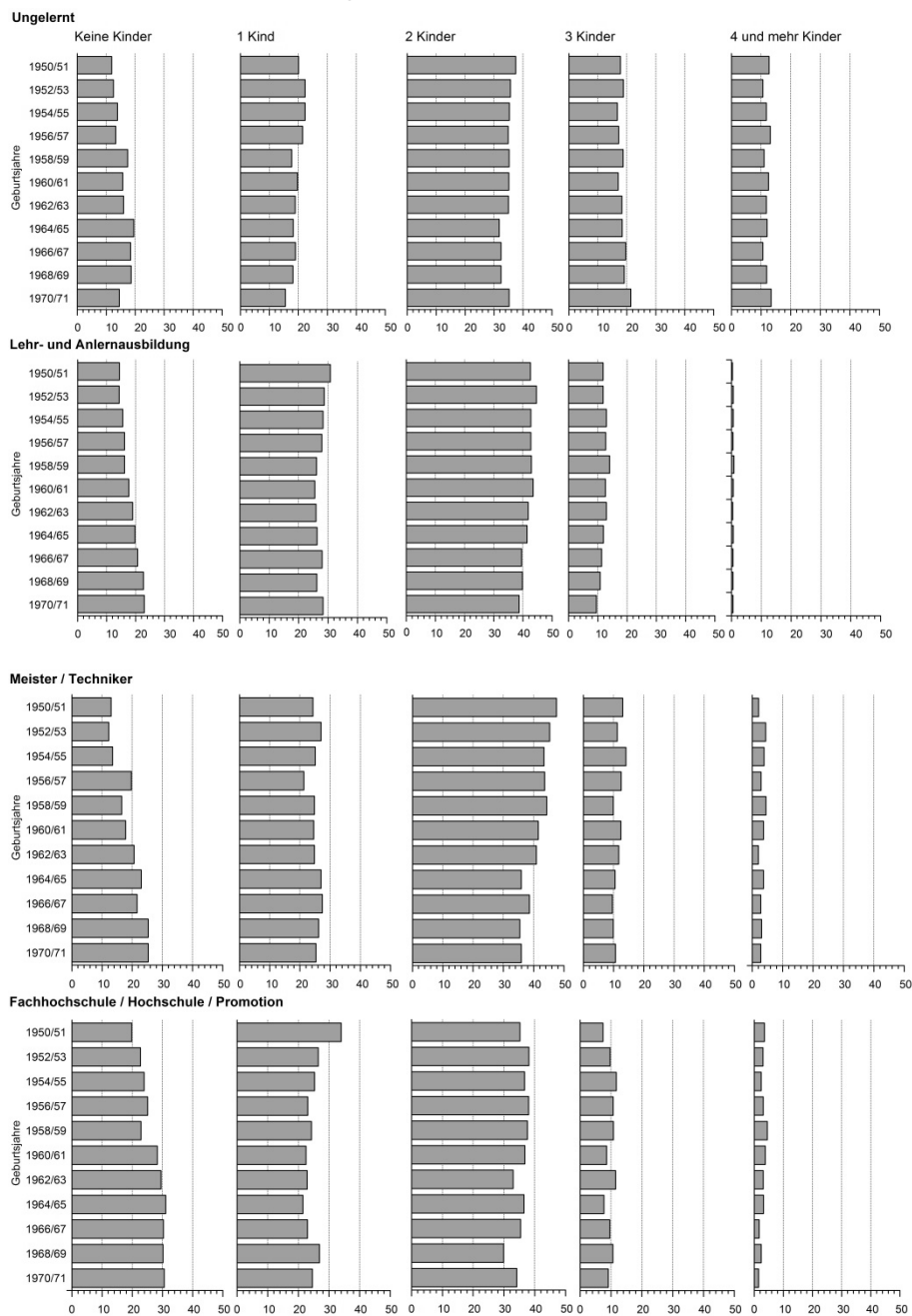
Auf einige Schwächen sei hingewiesen: In der Dendrogrammanalyse wurden veränderungsresistente und fluide Zustandsdaten kombiniert. Beispielsweise sind Ost/West und Bildung relativ stabile Merkmale, während Partnerschaft/-losigkeit relativ fluide ist. Zudem bleibt die Kausalitätsfrage offen: Haben Frauen in Großstädten weniger Kinder wegen eines urbanen Effektes oder aufgrund von Binnenmobilität kinderloser Frauen in Städte? Oder sind Vollzeit arbeitende Frauen häufiger kinderlos, weil sie in der Vergangenheit berufsorientierter waren, oder arbeiten Frauen, die kinderlos geblieben sind, einfach mehr? Für diese Fragen sind Ereignisanalysen anhand von Paneldaten wie SOEP, pairfam oder GGS unabdingbar und die Grenzen des MZ offensichtlich.

Die Subgruppenanalyse auch in hier nicht gezeigten Varianten verdeutlicht, dass die Effekte dieser fünf Merkmale sich erheblich ergänzen und in jeder Subgruppe nachweisbar sind. Es gibt keine Fallkombination, bei der die Effekte von Vollzeit, Partnerlosigkeit, hoher Bildung, Großstadt oder Westdeutschland *nicht* jeweils die Kinderlosigkeit erhöhen.

#### *4.2 Paritätsspezifische Differenzierung nach Kohorten und beruflicher Bildung*

In diesem Beispiel wird gezeigt, wie sich die Dimensionen Paritäten, Sozialstruktur und Jahrgänge kombinieren lassen. Diese weitgliedrige Strukturierung der MZ-Daten differenziert Frauen nach Geburtsjahrgängen, der Kinderzahl und der beruflichen Bildung (Abb. 7). Dabei wird zwischen elf Geburtsjahrgangsgruppen, fünf Paritäten und vier beruflichen Bildungsabschlüssen unterschieden. Ergebnis ist eine spezifische Sicht auf den Wandel der Fertilitätsmuster in Abhängigkeit von der Bildung.

**Abbildung 7:** Anteil Frauen nach Zahl der geborenen Kinder, Geburtsjahrgang und beruflicher Bildung in Deutschland



Quelle: Eigene Berechnung und Darstellung basierend auf MZ 2012.

Die Paritätsverteilungen der Geburtsjahrgänge 1950/51 bis 1968/69 zeigen Konstanz und Wandel. Gemeinsam für alle beruflichen Bildungsabschlüsse ist erstens, dass die Frauen am häufigsten zwei Kinder zur Welt gebracht haben. Die Anteile schwanken um 40%. Sie sind bei den Frauen mit einer Lehr- und Anlernausbildung<sup>13</sup> und den Meisterinnen und Technikerinnen etwas höher als bei Ungelernten und den Frauen mit einem Fachhochschul- oder Hochschulabschluss bzw. einer Promotion. Zweitens ist bei allen Bildungsstufen die Kinderlosigkeit in den Jahrgängen 1968/69 höher als in den Jahrgängen 1950/51. Die Kinderlosigkeit steigt mit jeder Bildungsstufe an. Mit ca. 30% weisen die hochgebildeten Frauen der Geburtsjahrgänge 1964/65 die höchsten Kinderlosenanteile auf, seitdem ist hier der Anstieg der Kinderlosigkeit gestoppt (vgl. Bujard 2012; Statistisches Bundesamt 2013a). Bedeutsam ist, dass in allen beruflichen Bildungsstufen der Anstieg der Kinderlosigkeit aufgehört bzw. sich deutlich verlangsamt hat. Drittens ist Kinderreichtum mit Ausnahme der Frauen ohne Berufsabschluss selten geworden. Hier sind die Frauen des Jahrgangs 1968/69 zu über 30% kinderreich. Es zeichnet sich sogar ein Wiederanstiegstrend ab.

Trotz des Wandels in den Paritätsmustern sind typische bildungsspezifische Unterschiede erhalten geblieben. Frauen ohne Abschluss sind seltener kinderlos und haben weniger ein oder zwei Kinder, dagegen sind die Anteile der Frauen mit drei oder vier Kindern höher als in den anderen beruflichen Bildungsgruppen. Die Anteile der Frauen mit nur einem Kind sind in dieser Gruppe am niedrigsten. Frauen, die promoviert sind oder eine Fachhochschule oder Hochschule absolviert haben, waren schon immer häufiger kinderlos und sind dies auch geblieben. Frauen, die über eine Lehr- oder Anlernausbildung verfügen oder Meister oder Techniker sind, weisen ähnliche Paritätsverteilungen auf, die zwischen den Mustern der Niedrig- bzw. Hochqualifizierten rangieren. Die Bildungsgradienten sind unterschiedlich: Bei der Kinderlosigkeit ist er linear, bei den Frauen mit vier oder mehr Kindern u-förmig.

Es ist generell beachtenswert, wie stabil die Fertilitätsmuster über elf Jahrgangsgruppen geblieben sind und sich Spezifika erhalten haben. Dies ist im Kontext der Diskussion um das zukünftige Fertilitätsniveau in Deutschland bedeutsam. Voraussetzung für einen nachhaltigen Fertilitätsanstieg ist ein Wandel in den Paritätsstrukturen, getragen von einem Rückgang bei der Parität 0 und Zuwächsen bei den Paritäten 2 und 3. Ein solcher Wandel hat noch nicht grundsätzlich eingesetzt, ist aber in Ansätzen zu erkennen. In der Gruppe der Ungelernten sind ein Rückgang der Kinderlosigkeit und ein Anstieg bei den Anteilen dritter sowie vierter und weiterer Kinder zu erkennen. Generell scheint der Anstieg der Kinderlosigkeit in allen beruflichen Ausbildungsgruppen gestoppt zu sein. Diese Erkenntnis ist vor allem bei den Hochqualifizierten, die maßgeblich zum hohen Niveau der Kinderlosigkeit in Deutschland beitragen, von Bedeutung.

---

13 Die Zusammenfassung von Lehr- und Anlernausbildung ist nicht unproblematisch, da letztere nach dem Berufsbildungsgesetz von 1969 nicht als berufsqualifizierender Abschluss gilt. Da weder eine Ausklammerung dieser Gruppe noch eine andere Zuordnung als sinnvoll erscheint, wurde diese Klassifizierung vorgenommen. Der Anteil der Frauen mit Anlernausbildung beträgt nur 1,2% dieser zusammengefassten Gruppe.



#### *4.3 Zeitliche Differenzierung im Kohortenverlauf: Entwicklung des Erstgebäralters nach Bildungsgruppen*

Das dritte Beispiel zeigt, wie auf Basis sämtlicher bisheriger 39 Mikrozensus der Jahre 1972-2013<sup>14</sup> Entwicklungsprozesse einzelner Kohorten differenziert nach den jeweiligen Altersjahren der Frauen abgebildet werden können. Auf Basis der MZ 1972-2013 und der Informationen zu Frauen mit unter einjährigen Kindern im Haushalt wird im Folgenden das Alter der Erstgebärenden in Westdeutschland analysiert. Dies wird für alle Altersjahre von 18 bis 49 für die Kohorten 1954, 1964, 1974 und 1979 differenziert nach Akademikerinnen und Nichtakademikerinnen vorgenommen. Da nach dem 44. Geburtstag fast keine Erstgeburten stattfinden und die bildungsspezifischen Werte vor dem Alter 25 nicht sinnvoll interpretierbar sind, werden die Erstgeburten für die Altersgruppe 25-45 gezeigt (Abb. 8). Während die Fallzahlen bei den Nichtakademikerinnen etwa zwischen 3.000 und 5.000 pro Geburtsjahr der Frau und Alter liegen (also pro Punkt der Abbildung), sind sie bei Akademikerinnen meistens in der Spanne von etwa 200-600. Daher wurden die Werte für Akademikerinnen durch den Drei-Jahres-Durchschnitt geglättet.

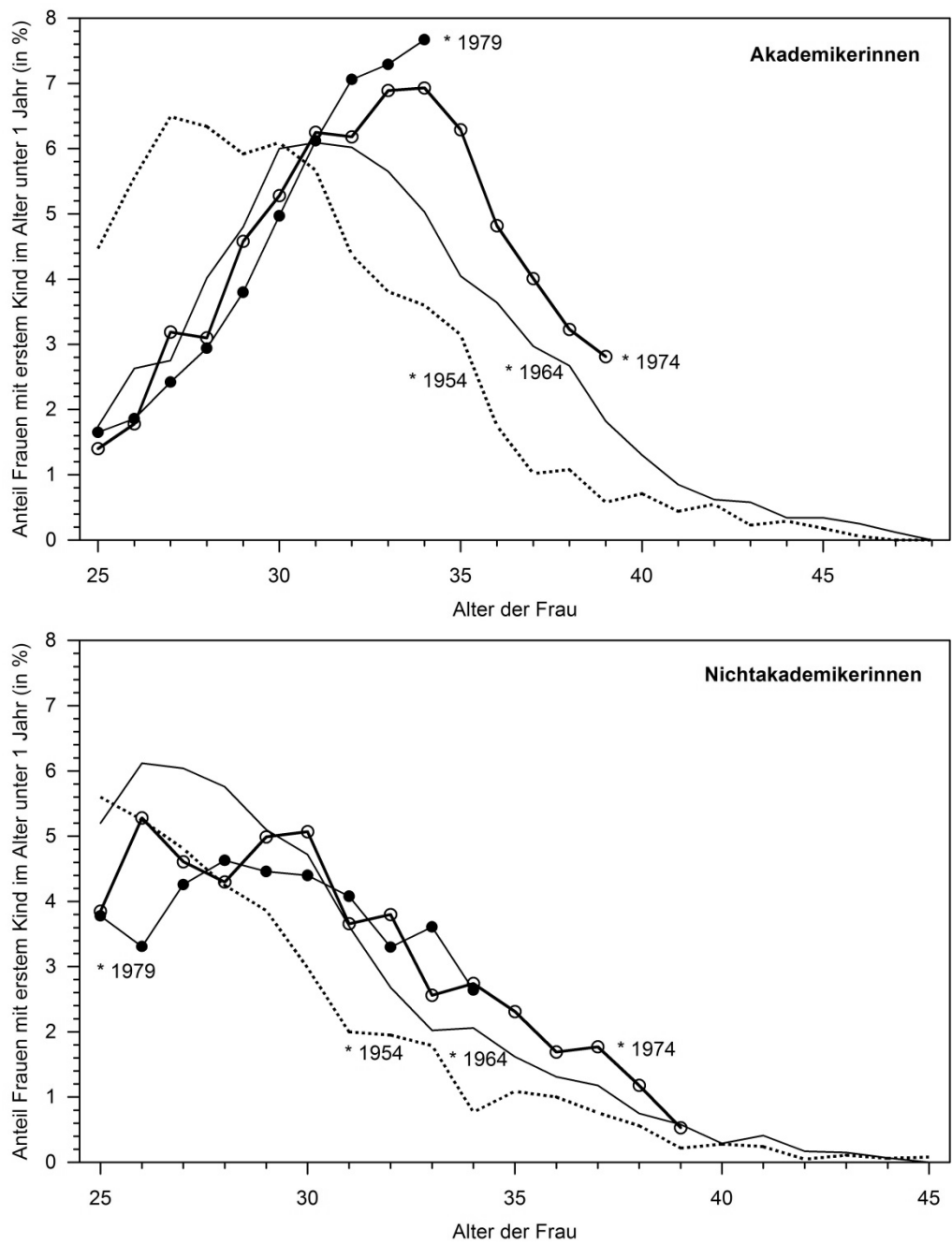
Die Verläufe verdeutlichen den Anstieg des Alters bei der Erstgeburt durch die Verschiebung der Kurven nach rechts. Die zunehmende Zahl an Erstgeburten nach dem 30. Geburtstag kompensiert früher aufgeschobene Geburten, jedoch zeigt die Analyse auch, dass sich ab Mitte 30 der Recuperationprozess aufgrund biologischer Grenzen abschwächt (vgl. Pötzsch 2013). Bei Akademikerinnen zeigt sich die Komprimierung der Geburten in der Rushhour der Lebensentscheidungen darin, dass der Anteil von Erstgeburten bis zu 7,7% in einem Altersjahr ausmacht (34-Jährige der 1979er Kohorte). Im Vergleich: Der höchste Anteil dieser Kohorte bei den Nichtakademikerinnen liegt bei 4,6% im Alter von 28 Jahren.

Bemerkenswert ist der sprunghafte Anstieg der Erstgeburten bei Akademikerinnen im Alter von 33-39 Jahren, insbesondere in den jüngeren Kohorten. Hier findet – spät im Lebensverlauf – ein intensiver Prozess nachholender Geburten statt. Der Sprung bei der Kohorte 1974 mit 33 und 34 Jahren entspricht den Jahren 2007 und 2008. Da das Elterngeld zu Beginn 2007 eingeführt wurde, könnte dies mit dem Befund von Bujard und Passet (2013) zusammenhängen, wonach das Elterngeld bei den Akademikerinnen einen Effekt aufgewiesen hat. Der Vergleich des Erstgebäralters nach Bildung, der nur mit dem MZ möglich ist, verdeutlicht, wie unterschiedlich die Verläufe im Lebensverlauf zwischen Akademikerinnen und den anderen Frauen sind.

---

14 Für 1975, 1983 und 1984 liegen keine MZ vor. Fehlende Werte wurden durch die Abgrenzung von Kindern zwischen ein und zwei Jahren des folgenden MZ vorgenommen.

Abbildung 8: Anteil von Erstgeburten nach Bildungsabschluss und Alter für die Geburtsjahre 1954-1979 in Westdeutschland



Quelle: Datenbasis MZ 1972-2013, eigene Berechnung und Darstellung.

## 5. Zusammenfassung und Diskussion

Der Beitrag diskutiert die Stärken und Limitationen des MZ anhand mehrerer empirischer Beispiele zu demografischen Fragestellungen. Aufgrund der hohen Fallzahlen – 680.563 Personen am Hauptwohnsitz im MZ 2012 – und dem verpflichtenden Charakter der Befragung sind die Potenziale für demografische und familiensoziologische Fragestellungen enorm. Anhand von Konfidenzintervallen im Vergleich von MZ und SOEP wurde die Präzision von gruppenspezifischen Häufigkeiten systematisch diskutiert. Die hier gezeigten Analysen gelten grundsätzlich für alle MZ und für verschiedene Fragestellungen. Bezüglich demografischer Fragen wie der Kinderlosigkeit gelten sie insbesondere für die MZ 2008 und 2012 mit der Geburtenfrage.

Diese Analysepotenziale lassen sich durch Forschungsdesigns heben, die die Stichprobengröße nutzen, indem sie Verteilungen und Verläufe der drei Dimensionen Sozialstruktur, Paritäten und Zeit tiefenscharf sichtbar machen. Dies wird durch drei Beispiele demonstriert:

- Differenzierungen der Sozialstruktur: Anhand von Dendrogrammanalysen wurde für Frauen der Kohorten 1950-69 die Kinderlosigkeitsrate nach fünf Ebenen und 48 Subgruppen differenziert. In diesen liegen die Kinderlosigkeitsraten zwischen 3,5% und 62,7%.
- Differenzierungen nach Paritäten: Die kombinierte Darstellung von Paritäten, Jahrgängen und beruflicher Bildung verdeutlicht, dass die Kinderlosigkeit bei allen Bildungsgruppen außer den Akademikerinnen ansteigt und der Anteil kinderreicher Frauen konstant verläuft.
- Differenzierungen nach Zeit bzw. Frauenjahrgängen: Die dazu gezeigten Analysen kombinieren Kohorten und Altersjahre basierend auf den 39 Mikrozensus zwischen 1972 und 2013. Sie zeigen im Kohortenvergleich, wie sehr das Erstgebäralter bei Akademikerinnen ansteigt.

Der MZ hat in mehrerer Hinsicht völlig andere Charakteristika im Vergleich zu anderen sozialwissenschaftlichen Datensätzen wie SOEP, NEPS, pairfam, AID:A, FiD oder GGS. Er hat ihnen gegenüber den Nachteil, dass er mit der Geburtenfrage kein Paneldesign hat und dadurch Methoden wie Ereignisanalysen und entsprechende Kausalitätsbefunde nicht möglich sind. Andererseits hat er den Vorteil, dass die Fallzahlen im Querschnitt 32-fach höher sind als beim SOEP, 50-fach höher als bei pairfam und 161-fach höher als beim GGS der zweiten Welle. Der Vergleich von gruppenspezifischen Werten zur Kinderlosigkeit zwischen SOEP und MZ hat gezeigt, dass die Werte dann nur weniger als einen Prozentpunkt auseinander liegen, wenn die Fallzahl der Gruppe auch im SOEP über 1000 liegt. Bei  $N < 500$  im SOEP waren Differenzen von mehreren Prozentpunkten zu finden.

Diese Präzisionsunterschiede wurden anhand von Konfidenzintervallen belegt: Werte zur Kinderlosigkeit der einzelnen Jahrgänge 1960-1971 haben im MZ 95%-ige Konfidenzintervalle von 1,1 Prozentpunkten nach oben und unten, während die Schwankung beim SOEP das Fünffache beträgt. Durch das hohe N im MZ lassen sich demografische Strukturen und Prozesse in einer Differenziertheit analysieren, die mit anderen Datensätzen (Zensus 2011 ausgenommen) nicht möglich sind. Stärken und Schwächen von MZ und den genannten Panels sind daher komplementär und die sozialwissenschaftliche For-

sung profitiert, wenn beide Informationsquellen mit entsprechenden Forschungsdesigns und Methoden ausgeschöpft werden.

Die Grenzen des MZ liegen v.a. im Querschnittsdesign und darin, dass Fertilitätsdaten zu Männern, zu Binnenmobilität und zu bilokalen Paarbeziehungen nicht vorliegen. Insbesondere eine Erhebung der leiblichen Kinder von Männern auf freiwilliger Basis wäre eine bedeutende Weiterentwicklung des MZ (vgl. Stock et al. 2012). Denn der Forschungsstand zu Kinderlosigkeit und CTFR von Männern ist noch geringer als bei den Frauen vor den MZ 2008 und 2012, da die Informationen zu den Kindern im Haushalt bei den Männern kaum sinnvolle Schätzmodelle erlauben, denn es gibt weitaus mehr Väter als Mütter, die nicht mit ihren Kindern im Haushalt leben.

Einige methodische Herausforderungen wurden diskutiert: Das Problem der Jahresüberhänge lässt sich durch Imputationen lösen. Da die Antwort auf die Geburtenfrage freiwillig ist, sind die Antwortverweigerer eine potenzielle Quelle von Verzerrungen. Es wird gezeigt, dass die Art der Erhebung einen immensen Einfluss auf die Beantwortung der Geburtenfrage hat: Hier sind Verzerrungen durch eine Unterschätzung der Kinderzahl möglich, wenn Fremdauskunft oder Selbstausfüller vorliegt, da bei beidem die Kinderlosigkeit und die Non-Response-Rate höher sind. Diese Verzerrung wurde durch die Imputation deutlich reduziert. Die gesamte Non-Response-Rate liegt bei 8,4% und nach den hier diskutierten Imputationen bei 3,7%. Gegenüber dem MZ 2008 ist aber die Verlässlichkeit im MZ 2012 noch weiter gestiegen, aus Lerneffekten und da die Geburtenfrage im Fragebogen in einem passenderen Kontext platziert wurde.

Zweifellos haben die MZ 2008 und 2012 erstmals in Deutschland verlässliche Daten zur Kinderlosigkeit und zur bildungsspezifischen Kohortenfertilität zur Verfügung gestellt, die auch einschlägig publiziert worden sind (Kreyenfeld/Konietzka 2013; Statistisches Bundesamt 2009a; 2013a). Hervorzuheben ist, dass das 1996 eingeführte Konzept der Lebensformen mit den Angaben zur Geburtenfolge der Kinder untersetzt werden kann, so dass besondere Rückschlüsse auf das generative Verhalten von verheirateten Paaren, nichtehelichen Lebensgemeinschaften oder Alleinlebenden gezogen werden können.

Allerdings sind einige der Anwendungsmöglichkeiten noch nicht genutzt worden. Abgesehen von logistischen Regressionen, die beim MZ auch für sehr spezialisierte Fragestellungen möglich sind (Bujard *in diesem Band*; Naderi *in diesem Band*; siehe auch: Kreyenfeld et al. 2009), lassen sich anhand deskriptiver statistischer Methoden weitreichende Differenzierungen der drei Dimensionen Sozialstruktur, Paritäten und Zeit – und ihrer Kombination – generieren. Für die Sozialwissenschaft ergibt sich durch den MZ ein großer Fundus an demografischen und sozialstrukturellen Analysemöglichkeiten. Allerdings benötigt man adäquate Forschungsdesigns und eine angemessene Interpretation der Querschnittsdaten.

### *Danksagung*

Wir danken Samira Beringer, Evelyn Grünheid, Detlev Lück, Robert Naderi, Norbert F. Schneider, Stine Waibel sowie zwei anonymen Gutachter(inne)n für wertvolle Anmerkungen.

## Literatur

- Birg, H. (2003). *Die demographische Zeitenwende*. München: Beck.
- Birg, H. & Flöthmann E.-J. (1996). *Entwicklung der Familienstrukturen und ihre Auswirkungen auf die Belastungs- bzw. Transferquotienten zwischen den Generationen*. Bielefeld: Institut für Bevölkerungsforschung und Sozialpolitik der Universität Bielefeld (IBS).
- Bujard, M. (2012). *Talsole bei Akademikerinnen durchschritten? Kinderzahl und Kinderlosigkeit in Deutschland nach Bildungs- und Berufsgruppen*. Wiesbaden: Bundesinstitut für Bevölkerungsforschung (BiB Working Paper 4/2012).
- Bujard, M. & Passet, J. (2013). Wirkungen des Elterngelds auf Einkommen und Fertilität. *Zeitschrift für Familienforschung/Journal of Family Research* 25, 2, S. 212-237.
- Deutscher Bundestag (2007). *Entwurf eines Gesetzes zur Änderung des Mikrozensusgesetzes 2005 und des Bevölkerungsstatistikgesetzes*. Berlin: Drucksache 16/5239.
- DIW & SOEP (Hrsg.) (2011). *Documentation PGEN. Person-related status and generated variables*. Berlin: Deutsches Institut für Wirtschaftsforschung/Sozio-oekonomisches Panel.
- DIW & SOEP (Hrsg.) (2013). *SOEP 2012 – Methodenbericht zum Befragungsjahr 2012 (Welle 29)*. Berlin: Deutsches Institut für Wirtschaftsforschung/Sozio-oekonomisches Panel (SOEP Survey Papers, 144).
- Dorbritz, J. & Ruckdeschel, K. (2013). Kinderlosigkeit – differenzierte Analysen und europäische Vergleiche. In: Kreyenfeld, M. & Konietzka, D. (Hrsg.), *Ein Leben ohne Kinder*. Wiesbaden: Springer VS, S. 253-278.
- Dreschmitt, K. & Naderi, R. (2015). *Sozialwissenschaftliche Daten zur Erforschung von Kinderlosigkeit und Kinderreichtum*. Wiesbaden: Bundesinstitut für Bevölkerungsforschung (BiB Working Paper 04/2015).
- Duschek, K. & Wirth, H. (2005). Kinderlosigkeit von Frauen im Spiegel des Mikrozensus. *Wirtschaft und Statistik*, 8/2005, S. 800-820.
- Frick, J. & Schmitt, C. (2013). BIOBIRTH – Data Set on the birth biography of female respondents. In: *SOEP Group 2013. SOEP 2012 – Documentation on biography and life history data for SOEP v29*. Berlin: Deutsches Institut für Wirtschaftsforschung/Sozio-oekonomisches Panel (SOEP Survey Papers 176).
- Häder, M. & Häder, S. (2014). Stichprobenziehung in der quantitativen Sozialforschung. In: Baur, N. & Blasius, J. (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS, S. 283-297.
- Kohler, H.-P., Billari, F. & Ortega, J. (2002). The emergence of lowest-low fertility in Europe during the 1990s. *Population and Development Review*, 28, S. 641-680.
- Kreyenfeld, M. & Konietzka, D. (2013). Kinderlosigkeit in Deutschland. In: Kreyenfeld, M. & Konietzka, D. (Hrsg.), *Ein Leben ohne Kinder*. Wiesbaden: Springer VS, S. 13-44.
- Kreyenfeld, M., Schmidtke, K. & Zühlke, S. (2009). Eignet sich das Mikrozensus-Panel für familiensoziologische Fragestellungen? *Zeitschrift für Familienforschung/Journal of Family Research*, 21, 3, S. 264-285.
- Kreyenfeld, M., Zeman, K., Burkimsher, M. & Jaschinski, I. (2011). Fertility data for German speaking countries. *Comparative Population Studies*, 37, S. 349-380.
- Kroh, M. (2014). *Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP) (1984 until 2012)*. Berlin: Deutsches Institut für Wirtschaftsforschung/Sozio-oekonomisches Panel (SOEP Survey Papers 177).
- Krupp, H.-J. (2008). Die Anfänge: Zur Entstehungsgeschichte des SOEP. *Vierteljahreshefte zur Wirtschaftsforschung*, 77, 3, S. 15-26.
- Kühnel, S. & Krebs, D. (2010). Grundlagen des statistischen Schließens. In: Wolf, C. & Best, H. (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: Springer VS, S. 165-189.
- Lengerer, A., Janßen, A. & Bohr, J. (2007). Familiensoziologische Analysepotenziale des Mikrozensus. *Zeitschrift für Familienforschung*, 19, S. 186-209.

- Miettinen, A., Rotkirch, A., Szalma, I., Donno, A. & Tanturri, M. L. (2015). Increasing childlessness in Europe: Time trends and country differences. *Families and Societies Working Paper* 33.
- OECD (2015). Family Database. *SF 2.5 Childlessness*. [www.oecd.org/els/family/database.htm](http://www.oecd.org/els/family/database.htm) [Stand: 2015.05.05].
- Pötzsch, O. (2013). Wie wirkt sich der Geburtenaufschub auf die Kohortenfertilität in West und Ost aus? *Wirtschaft und Statistik*, 2/2013, S. 87-101.
- SAS Institute Inc. (2008): *SAS/STAT® 9.2 User's guide*. Corry: SAS Institute Inc.
- Scharein, M. & Unger, R. (2005). Kinderlosigkeit bei Akademikerinnen? *BIB-Mitteilungen* 26, 2, S. 6-13.
- Schimpl-Neimanns, B. (2011). Schätzung des Stichprobenfehlers in Mikrozensus Scientific Use Files ab 2005. *AStA: Wirtschafts- und Sozialstatistisches Archiv*, 5, S. 19-38.
- Schimpl-Neimanns, B. (2013). Methodische Herausforderungen bei der Erfassung von Bildung und Ausbildung im Mikrozensus. In: Riede, T., Bechtold, S. & Ott, N. (Hrsg.): *Weiterentwicklung der amtlichen Haushaltsstatistiken*. Berlin: Scivero, S. 43-80.
- Schmitt, C. & Wagner, G. (2006). Kinderlosigkeit von Akademikerinnen überbewertet. *Wochenbericht DIW*, Nr. 21/2006, S. 313-317.
- Schneider, N. F. (1996): Bewußt kinderlose Ehepaare. *Zeitschrift für Frauenforschung*, 14, S. 128-137.
- Schupp, J., Schwarze, J. & Wagner, G. (1997). Erwerbsstatistik unterschätzt Beschäftigung um 2 Millionen Personen. *Wochenbericht DIW*, Nr. 38/1997, S. 689-696.
- Sobotka, T. (2005). *Childless societies? Trends and projections of childlessness in Europe and the Unites States*. Paper presented at the 2005 Population Association of America (PAA) Meeting, Philadelphia.
- Statistische Ämter des Bundes und der Länder (o. J.). *Informationen zur Durchführung der Erhebung 2012*. Statistisches Bundesamt (2006). *Kinderlosigkeit von Akademikerinnen im Spiegel des Mikrozensus*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2009a). *Mikrozensus 2008. Neue Daten zur Kinderlosigkeit in Deutschland*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2009b). Methodeninformation: Imputation von Werten bei fehlenden Angaben zur Mutterschaft und zur Zahl der geborenen Kinder im Mikrozensus 2008. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2013a). *Geburtentrends und Familiensituation in Deutschland*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2013b). *Mikrozensus 2012. Qualitätsbericht*. Wiesbaden: Statistisches Bundesamt.
- Stock, G., Bertram, H., Fürnkranz-Prskawetz, A., Holzgreve, W., Kohli, M. & Staudinger, U. (Hrsg.) (2012). *Zukunft mit Kindern*. Frankfurt am Main: Campus.
- Wagner, G.; Frick, J. & Schupp, J. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, evolution and enhancement. *Schmollers Jahrbuch*, 127, 1, S. 139-169.
- Wirth, H. (2006). *Die kinderlosen Akademikerinnen*. Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA).
- Wirth H. & Dümmler, K. (2004). Zunehmende Tendenz zu späteren Geburten und Kinderlosigkeit bei Akademikerinnen. *Informationsdienst Soziale Indikatoren*, 32, S. 1-6.
- Wirth H. & Dümmler, K. (2005). Der Einfluss der Qualifikation auf die Kinderlosigkeit von Frauen zwischen 1970 und 2001 in Westdeutschland. *Zeitschrift für Bevölkerungswissenschaft*, 30, 2-3, S. 313-336.
- Wolf, C. & Lüttinger, P. (2009). Verteilung von Proxy-Interviews im deutschen Mikrozensus. In: Weichbold, M., Bacher, J. & Wolf, C. (Hrsg.), *Umfrageforschung. Herausforderungen und Grenzen*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 395-412.

Eingereicht am/Submitted on: 24.07.2015

Angenommen am/Accepted on: 12.11.2015

## Anschriften der Autoren und der Autorin/Addresses of the authors:

Dr. Martin Bujard  
Dr. Jürgen Dorbritz  
Linda Lux, Diplom-Soziologin

Bundesinstitut für Bevölkerungsforschung (BiB)/  
Federal Institute for Population Research  
Friedrich-Ebert-Allee 4  
65195 Wiesbaden  
Deutschland/Germany

Dr. Robert Herter-Eschweiler  
Statistisches Bundesamt (Destatis)/  
Federal Statistical Office  
Graurheindorfer Str. 198  
53117 Bonn  
Deutschland/Germany

E-Mail: martin.bujard@bib.bund.de  
juergen.dorbritz@bib.bund.de  
robert.herter-eschweiler@destatis.de  
linda.lux@bib.bund.de